THE UNIVERSITY
of EDINBURGH

A New Model for Long-range Chromatin Reorganisation upon Enhancer-driven Gene Activation.

THE UNIVERSITY *of* EDINBURGH

Nezha Suzanne BENABDALLAH

PhD by Research

University of Edinburgh 2016

Table of Contents

List of Figures

Chapter 1: Introduction

Chapter 2: Shh-Brain-Enhancer 6 (SBE6), a novel long-range enhancer driving Shh expression in neural progenitor cells.

Chapter 3: Analysing Shh promoter-enhancer interactions using imaging and 5C

Chapter 4: Using a synthetic biology approach to study enhancer-promoter interaction

Chapter 5: Analysing the mechanism of the chromatin unfolding

Chapter 6: Shh regulatory region insulation

Chapter 7: Discussion

Chapter 8: Material & Methods

List of Tables

Declaration

I declare that this thesis has been entirely written by me and is my own work. The work has not been submitted for any other degree or professional qualification.

Data and ideas arising from this thesis have resulted in article and review publications. Ideas discussed in the introduction can be linked with (Benabdallah and Bickmore, 2015). Data and conclusions from Chapter 2 have resulted in a prepared manuscript (Benabdallah et al., 2016 *submitted*) and parts of data from Chapter 3 to 5 have resulted in another manuscript in preparation.

Nezha Benabdallah

Acknowlegments

I would like to thank first my supervisor Pr. Wendy Bickmore for the continuous support during my thesis and for giving me the chance to complete my PhD in her inspiring laboratory. Wendy gave me the fabulous opportunity of driving independently my own research and to follow my own goals. With her motivation, support and immense knowledge, my thesis took an exciting path using cutting-edge technology and synthetic biology tools to accomplish a step forward in the enhancer field.

I would also like to thank my second supervisor Paul Dalgarno, Doug Vernimmen and Pierre Therizols for their insightful comments and encouragement, but also for the hard question that widen my research from various perspectives.

I thank the whole lab and a special mention to Shelagh Boyle for scientific insight but also keeping the spirit of the lab alive, useful coffee breaks and amazing cake baking skills, Rob Illingworth for the scientific input but also keeping the spirit of the office alive and working crazy hours (it is always enjoyable to have company) and Iain Williamson for the support in the lab and the help with manuscripts and thesis writing.

In more personal acknowledgments I would like to thank Genevieve Almouzni who helped me forging my scientific career and strongly recommended me to join Wendy's lab. Also Mr. Anselm, my biology teacher in *classes preparatoires* and Mr. Olivieri my histoire-géographie teacher in middleschool, for teasing my curiosity and building my critical mind. They both had a crucial role in shaping my path and I am very grateful to have met them.

Lay Summary of Thesis

My thesis aims to contribute to the understanding of how gene expression is regulated. Indeed, a gene can be in an on or off state. A particular interest of mine is the study of DNA sequences (enhancers) that orchestrate the regulation of genes. Enhancers have to ability to switch on genes located far away in the chromosome. As genes controlled by enhancers are often critical for development and lead to severe defects when their activation is defective, it is important to understand them precisely. However, how enhancers work is still unclear. Knowing that enhancers are often located far from their target genes, a general assumption is that their mechanisms of action involves a long-range restructuring of chromatin. Nowadays, the established model for distal enhancers implies direct enhancer/promoter interaction with a looping out of intervening chromatin.

Using super-resolution imaging, I could study nuclear ultra-structure down to the 100 nm range. I monitored long-range chromatin reorganisation upon enhancer-driven activation of the Sonic hedgehog gene (Shh). This revealed an increase in nuclear distance between Shh and Shh-Brain-Enhancers that is not compatible with the looping model. Furthermore, using a synthetic biology approach, I could reveal that Shh-Brain-Enhancers regulate a long-range chromatin unfolding to enhance Shh expression, which portrays a unique but new mechanism for distal enhancers.

Abstract

Enhancers are non-coding DNA sequences which are able to activate the expression of a gene in a specific tissue manner and at a precise stage during embryonic development. First identified almost 40 years ago, our growing understanding of enhancers has transformed the concept of gene regulation to recognise the key role of these sequences in the expression of many genes. Moreover, the identification of human diseases caused by genetic variation in non-coding enhancer elements highlights the importance of characterising enhancers in order to understand human disease. However, enhancers are often located far from the promoter they influence and the mechanisms through which enhancers govern gene expression remain unclear. The most widely accepted model for the action of distal enhancers involves the formation of a chromatin loop, in which the enhancer and promoter physically interact at the loop base. The kinetics or molecular basis for the formation of enhancer/promoter loops is unknown and it remains unclear whether this mechanism of enhancer communication is universal, or indeed whether it is the most pervasive. The aim of my PhD is to investigate further the mechanism of action of distal enhancers in the regulation of developmental genes.

Using chromatin profiling during the differentiation of embryonic stem cells to neural progenitor cells in order to see which Shh enhancer is active in neural progenitor cells (NPCs), I report the identification of a novel long-range enhancer for *Shh* - Shh-Brain-Enhancer-6 (SBE6) – that is located 100kb upstream of Shh and that is required for the proper induction of *Shh* expression during a neural differentiation programme. SBE6 enhances *Shh* expression during the differentiation of neural progenitor cells (NPCs) and is active in the brain of developing zebrafish and mouse embryos.

Next, using a super-resolution 3D-FISH based approach to study the enhancer-driven activation of the Sonic hedgehog gene (*Shh*) I have identified a novel mechanism of long-range enhancer regulation that is incompatible with the looping model. Instead, gene activation is associated with an increase in nuclear distance between *Shh* and Shh-Brain-Enhancers. Using a synthetic biology approach I have determined that the chromatin unfolding is regulated specifically by the Shh-Brain-Enhancer and is mediated by the recruitment of transcription factor SIX3 and Poly (ADP-Ribose) Polymerase 1. Chromatin decondensation upon gene activation has been observed previously in *Drosophila* polytene

chromosomes. I suggest an analogous decompaction is driven by Shh-Brain-Enhancer to promote the activation of *Shh* in mouse neural progenitor cells. This 'chromatin unfolding' model represents a new mechanism of long-range enhancer-promoter communication in addition to the looping and tracking models.

Chapter 1: Introduction

Genes with vital roles in embryonic development have a highly regulated spatial and temporal expression pattern. The primary determinant of this precise tissue specific gene expression is believed to be a subset of non-coding regions in the genome called enhancers. Enhancers can regulate the promoters of their target genes over very large distances (up to 1Mb) and can be located upstream, downstream or within the introns of the target gene.

The growing comprehension of enhancer elements led to a complete redefinition of our concept of genetic regulation. First defined over 30 years ago, the mechanisms through which enhancers interact with the promoters and regulate target expression remain poorly understood despite the widespread acceptance of their crucial role in gene regulation.

It is commonly thought that when an enhancer is located in the vicinity of the core promoter, activation signals spreads between the enhancer and its cognate promoter by tracking or linking of proteins, and that distal enhancers are only thought to interact via a direct interaction with the core promoter that is enabled by the formation of a stable loop. However, it is now clear that more than a linear sequence, an enhancer is part of a wider domain that has the ability to regulate gene expression of promoters located in the same subdivision of the genome in the 3D nucleus space.

The growing recognition of the crucial part of enhancer mutations in both Mendelian and complex human genetic diseases makes it imperative to fully understand enhancer and their mechanisms by which they regulate gene expression.

1.1 Cis-regulatory elements

During embryonic development, highly specialized tissues are formed via the differentiation of cells into different lineages. In order to provide a precise patterning, morphogens undergo tight spatially and temporally controlled expression. A subset of noncoding regions of the genome called enhancers are the key determinants of precise tissue-specific gene activation. These cis-regulatory elements promote gene activation in a position and orientation-independent manner by recruiting tissue-specific transcription factors, co-activators and RNA polymerase II which are components also found to bind to the core promoters of genes (Koch et al., 2011).

The main attributes of cis-regulatory sequences were elucidated from studies dissecting the first ever described enhancer e.g. the 72 bp tandem repeat of SV40 DNA, a sequence located more than 100 nucleotides upstream from the core early viral promoter. Hence, enhancer deletion has been linked with reduced gene expression (Benoist and Chambon, 1981; Gruss et al., 1981). In parallel, other equivalent sequences were identified in mammalian cells (Banerji et al., 1983; Conrad and Botchan, 1982), including Locus Control Regions (LCR) – which define regulatory domains that act as a group of enhancers. The first LCR element identified upstream of the human ε, γ, δ and β-globin genes showed all the characteristics of the SV40 enhancer including strong clustered DNaseI hypersensitive sites (DHS) and distinctive histone marks (Grosveld et al. 1987; reviewed in Li 2002). Enhancer action has a highly precise tissue-specific activity (Schirm et al., 1987) and can be associated with other diverse promoters. Notably, cis-regulatory sequences that work over substantial distances (>10 kb) have been described (Banerji et al., 1981; Mellon et al., 1981; Moreau et al., 1981).

Chromosomal rearrangements studies have revealed that most of the functional enhancer are positioned within a range of ~1 Mb 5' or 3' in *cis* of their cognate promoter (Symmons and Spitz, 2013). However, position and orientation do not constrain enhancer action as both enhancer reporter assays (Bhatia et al., 2015) and localised transposon hopping (Anderson et al., 2014; Symmons et al., 2014) highlight that enhancer that endogenously act at long-range are also efficient at short range. Interestingly, many developmental genes are often neighbouring to gene deserts. The absence of other adjacent genes could help ensure regulatory specificity between enhancer and promoter. For example, artificial integration of

the β-globin locus control region (LCR) into a gene-rich domain leads to an ectopic activation of the adjacent genes (Noordermeer et al., 2008).

One well studied developmental gene regulated by distal enhancers is the sonic hedgehog gene (*Shh*). *Shh* is adjacent to a large gene desert spanning nearly 900 kb which contains a number a tissue-specific enhancers regulating its various expression domains (Epstein et al., 1999; Jeong, 2003, 2006; Sagai et al., 2004, 2009) (Figure Ch1-1). *Shh* encodes for a secreted morphogen that imparts patterns of growth and identity to cells during many stages of embryonic development, including neural progenitors throughout ventral regions of the developing central nervous system (CNS) (Dessaud et al., 2008; McMahon et al., 2003). *Shh* expression in neuroectodermal lineages is driven by Shh-Brain-Enhancer (SBE2, SBE3, SBE4 and SBE5) that are located roughly 300 kb upstream the *Shh* promoter (Jeong, 2006) or 800 kb for SBE5 (Yao et al., 2016). In the same manner, the ZRS (zone of polarizing activity regulatory sequence) located about 1 Mb away drives *Shh* expression during limb development (Lettice et al., 2003) revealing that the ZRS is able to bypass both the gene in which it is located - *Lmbr1*- and a neighbouring gene - *Rnf32* - in order to activate specifically *Shh* (Figure Ch1-1). The above example reveals that enhancer-promoter communication must carry very specific mechanisms of action as enhancers can also be positioned with other intervening genes or even located within the introns of nearby genes (Noonan and McCallion, 2010).

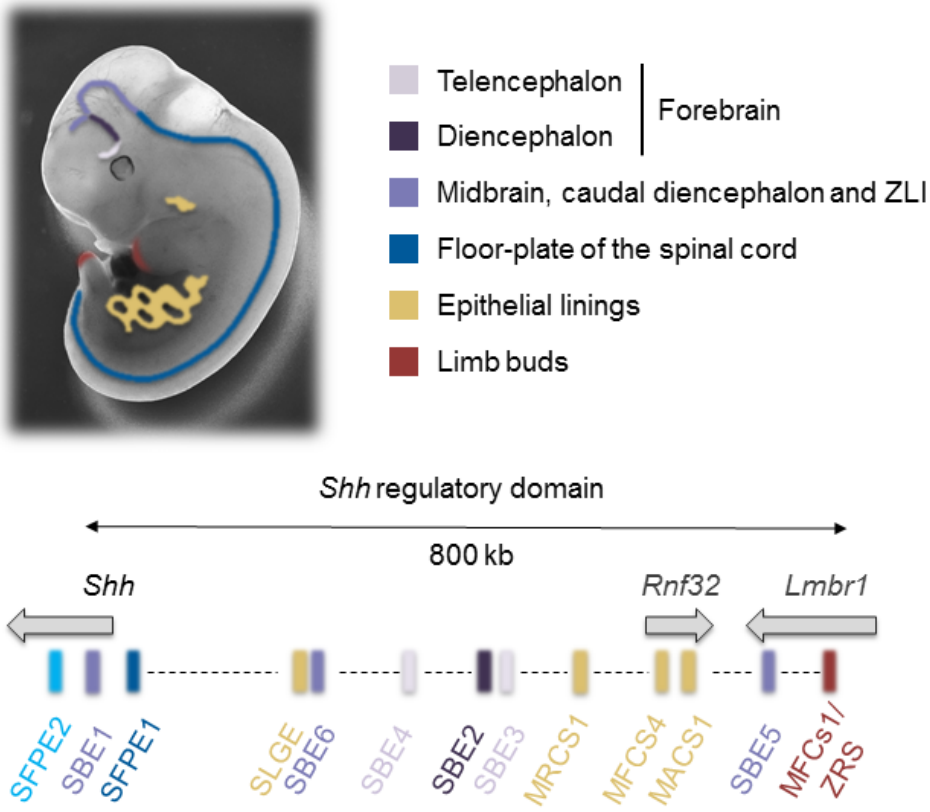

Figure Ch1- 1 Shh Promoter and tissue-specific enhancers.

Schematic (top) representing the activity of *Shh*-tissue-specific enhancers in mouse embryo and (bottom) map displaying the genomic locations of genes and enhancers in the *Shh* regulatory domain. Red box show ZRS, Shh limb bud enhancer, in blue are enhancers active in floor plate and spinal cord, in gold are enhancers for epithelial linings, and purple boxes represent enhancers driving *Shh* expression in the developing brain. Grey arrows represent direction of corresponding gene expression.

1.2 Enhancer-promoter communication

Over the years, several models have emerged regarding the communication between enhancers and promoters. Originally considered as an extension of the promoter, regulatory element were thought to provide a precise docking site for RNA polymerase II - or other components of the transcriptional machinery - followed by tracking of these factors on the chromatin fibre until they met and activated the cognate core promoter in a time and position-dependent manner (Ptashne and Gann, 1997).

The tracking model suggests a continuous spreading of the activation signal after proteins are recruited to the enhancer. Transcription factors could track along the chromatin fibre towards the promoter and associate with the polymerase to activate transcription or chromatin-remodelling factors spread the activation signal to the promoter (Kolesky et al., 2002). The linking model, which to date is sparsely defined, is similar to the tracking model, where the bound transcription factor leads to oligomerization of proteins that formed a bridge between then enhancer and the promoter (Bulger and Groudine, 1999, 2011; Mahmoudi, 2002) (Figure Ch1-2). Studies of spreading models at several loci, such as *β-globin*, have described an unidirectional spread of CBP/p300 histone acetyltransferases, histone acetylation (H3 and H4), components of chromatin remodelling complexes (SWI/SNF), or RNA polymerase II itself with synthesis of short polyadenylated, intergenic RNAs (Gribnau et al., 2000; Hatzis and Talianidis, 2002; Kim and Dean, 2004; Masternak et al., 2003; Spicuglia et al., 2002; Wang et al., 2005; Zhao and Dean, 2004; Zhu et al., 2007). These studies associate tracking as the primary step for enhancer-promoter interaction and hypothesize the later formation of a stable enhancer-promoter loop when tracking is complete (Figure Ch1-2).

The current widely accepted model for long-range communication implicates direct interactions between enhancer- and promoter-bound proteins that connects enhancer and promoter with looping out of the intervening chromatin (Su et al., 1991; Tolhuis et al., 2002) (Figure Ch1-2). This looping model may involve a transfer to the cognate promoter of RNA polymerase II or activating – or de-repressing – factors such as histone demethylase to remove Polycomb complex (Vernimmen et al., 2011).

Looping of the chromatin would enable a very specific enhancer-promoter communication, which illustrates the spatial and specificity of activation of a certain gene by a unique

enhancer. Recent elegant chromosome engineering experiments have demonstrated that imposing direct interaction between the *β-globin* promoter and its LCR is sufficient to generate *β-globin* activation in erythroid cells, where some other key transcription factors are also already bound (Deng et al., 2012, 2014). However, in some other cases enhancer-promoter direct interaction via chromatin looping may be insufficient for gene activation. It seems that chromatin looping for enhancer-promoter direct interaction can appear dynamic and cell-type specific (Benko et al., 2009; Simonis et al., 2006; Tolhuis et al., 2002) or more stable across cell types occurring in both expressing and non- expressing cells (de Laat and Duboule, 2013; Montavon et al., 2011) with transcription being initiated from stable pre-formed loops through paused polymerase release (Ghavi-Helm et al., 2014). Therefore enhancer-promoter loops can be stable across different tissues, even when the gene is not expressed, or more dynamic, with loop formation coinciding precisely with gene activation.

Considering chromatin as a polymer, limiting the diffusion of the fibre by the formation of a stable chromatin loop must have a substantial entropic cost. Therefore in theory we could predict the stability of a chromatin loop by knowing the size and the strength of binding-energies of enhancer and promoter-bound proteins. Likewise, for short-range interaction the rigidity of the chromatin fibre might prevent direct interaction via chromatin looping, as it would require incredibly strong protein interactions, confirming the convenience of the tracking model for short-range regulation (Vernimmen and Bickmore, 2015).

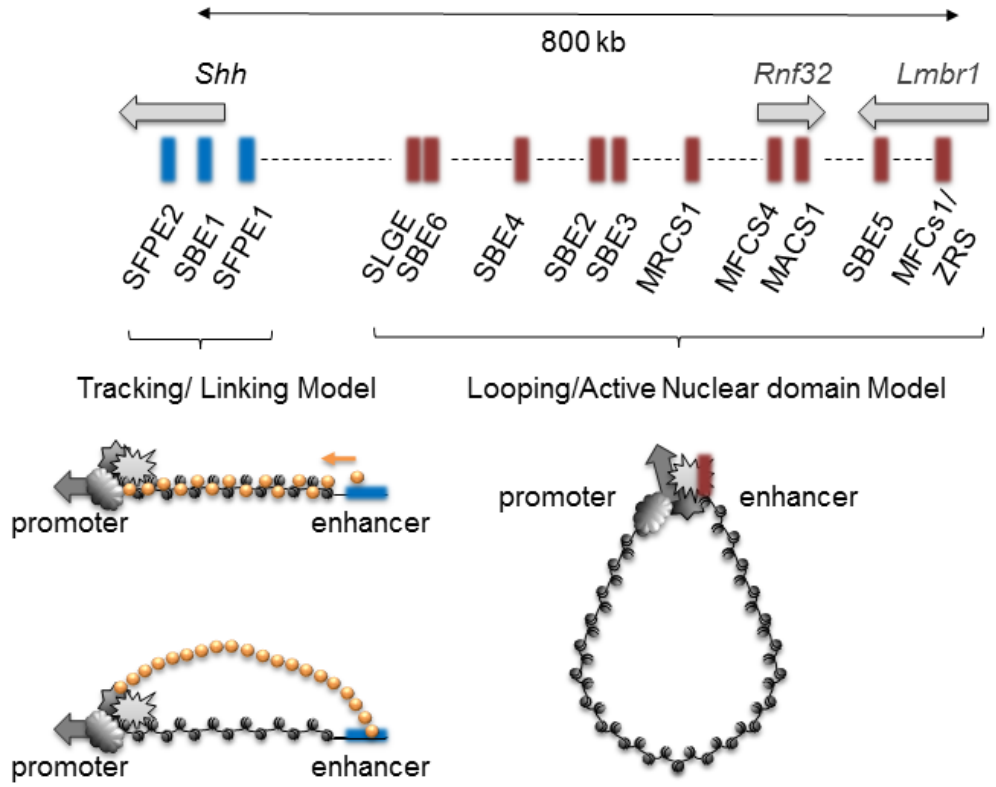

Figure Ch1- 2 Enhancer-promoter communication.

Enhancer and promoter communication mechanisms. Left - the tracking and linking models where neighbouring enhancer anchoring for activating signals that track along the chromatin fibre towards the cognate promoter (tracking model- up) or that oligomerize and form a protein bridge towards the cognate promoter (linking model - down) (Model not applicable for any *Shh* enhancer so far but is believed to be the case for proximal enhancer). Right - the looping model for distal enhancer where they directly interact with their cognate promoter by looping out of the intervening chromatin region (only shown for ZRS/Shh in that locus).

1.3 Chromatin organisation

Independently of the model for enhancer-promoter communication, it is clear that enhancers impact on genome regulation by modulating the chromatin architecture locally in a linear way or more broadly by changing the 3D folding of the locus.

1.3.1 A linear view of the chromatin state

In the interphase nucleus, chromatin is organized into functional domains with heterochromatin defined as a structure that does not change its condensation state during the cell cycle whereas euchromatin can appear decondensed and more thread-like during interphase (Augui et al., 2011; Hutchison and Weintraub, 1985; Zorn et al., 1976, 1979). Constitutive heterochromatin is localized mainly at the periphery of the nucleus and nucleolus (Croft et al., 1999; Kind and van Steensel, 2010), whereas euchromatin is distributed within the nucleoplasm and is associated with transcribed regions (Li et al., 2007; Lieberman-Aiden et al., 2009; Tanabe et al., 2002).

However, the nucleus is not a fixed entity and chromatin conformation changes contributing to the control of gene expression. Factors considered responsible for these different states are involved in the chromatin landscape (Filion et al., 2010). For example the chemical modifications of DNA such as methylation that occur without altering the sequence (Suzuki and Bird, 2008), the determinant of nucleosomes positioning and the post-translational histone modifications which influence DNA accessibility to activators and thus control gene expression (Kouzarides, 2007; Struhl and Segal, 2013), the binding of chromatin remodeller or transcription activators like Vp16 or acidic peptide that can decondense and rearrange large chromatin region (Belmont et al., 1999; Carpenter et al., 2005; Tumbar et al., 1999). Depending on the gene and the cell type, these determinants influence each other resulting in different chromatin state. This modularity has consequent effects on gene expression and represents a means to respond and adapt to external environmental stimuli.

More than just the local scale, modifications influencing chromatin structure changes can act over a very wide region of tens to hundreds kilobases (Bickmore and van Steensel, 2013; Cairns, 2009). Acetylation of histones H3 and H4 has been largely linked with open,

decondensed and thus active chromatin. A direct role for acetylation in chromatin decondensation has been established using incubation with the Histone Deacetylase (HDAC) inhibitor TSA (Eskeland et al., 2010a; Tóth et al., 2004). Indeed, acetylation reduces the positive electrostatic charge of the core histone and thus decreases the attraction of the negatively charged DNA to the histone core leading to a loose conformation and chromatin decondensation. In a general manner, acetylation occurs as one of the primary mechanisms for rapid change in gene expression as it can quickly rearrange chromatin upon a sudden stimulus such as *Drosophila's* polytene chromosomes puffs where rapid gene activation corresponds to a quick and massive chromatin decondensation displaying active marks such as H4K16ac (Turner et al., 1992).

However, the chromatin fibre is not just a further linear packaging of the DNA sequence than impacts its accessibility in 2D only. The chromatin can fold in more specific ways and can also adapt and change the overall 3D shape upon various determinants.

1.3.2 Chromatin domains and folding

1.3.2.1 Polycomb domains

Polycomb chromatin that can form large chromatin domains is described as a repressive type of chromatin that primarily regulates genes with developmental functions i.e facultative heterochromatin (Sparmann and van Lohuizen, 2006). Polycomb complexes PRC1 and PRC2 are well-known transcriptional regulators that have also been described as master chromatin architects by forming large domains covered by the H3K27me3 mark (Cao and Zhang, 2004; Ku et al., 2008; Pan et al., 2007). The canonical method by which this occurs depends on H3K27me3 deposition by PRC2 complex that is then recognised by and recruits units of the PRC1 complex to later induce chromatin compaction (Figure Ch1-3). The nucleation of a Polycomb complex further spread to cover a wider domain and leads to an expanded compaction of the whole domain. Chromatin architecture of some gene and enhancers clusters such as the HoxD locus are dictated by the Polycomb complexes (Eskeland et al., 2010b; Williamson et al., 2012, 2014).

Several studies have attempted to categorise chromatin based on the complement of histone modifications covering a particular locus and in each case a Polycomb repressed class of

chromatin has emerged as a distinct class (Filion et al., 2010; Ram et al., 2011; Zhu et al., 2013). Polycomb complexes are able to change nucleosomes occupancy and are responsible for chromatin compaction (Eskeland et al., 2010b; Francis et al., 2004; Ringrose and Paro, 2004). *In vitro* and *in vivo*, the Hox locus has been shown to decompact upon activation and to lose polycomb-complex and H3K27me3 marks (Chambeyron et al., 2005; Morey et al., 2007; Williamson et al., 2012). In some case, enhancers have been associated with active removal of polycomb marks on their target genes (Vernimmen et al., 2011). Interestingly, in *Drosophila*, it has been demonstrated that polycomb target genes co-localise and form defined domains called polycomb bodies (Bantignies et al., 2011; Hernandez-Munoz et al., 2005; Pirrotta and Li, 2012). In mammalian cells, there are also strong evidence for the formation of a 3D structure regrouping several Polycomb targets in a confine a nuclear space (Schoenfelder et al., 2015) (Illingworth, Boyle et al. in *preparation*).

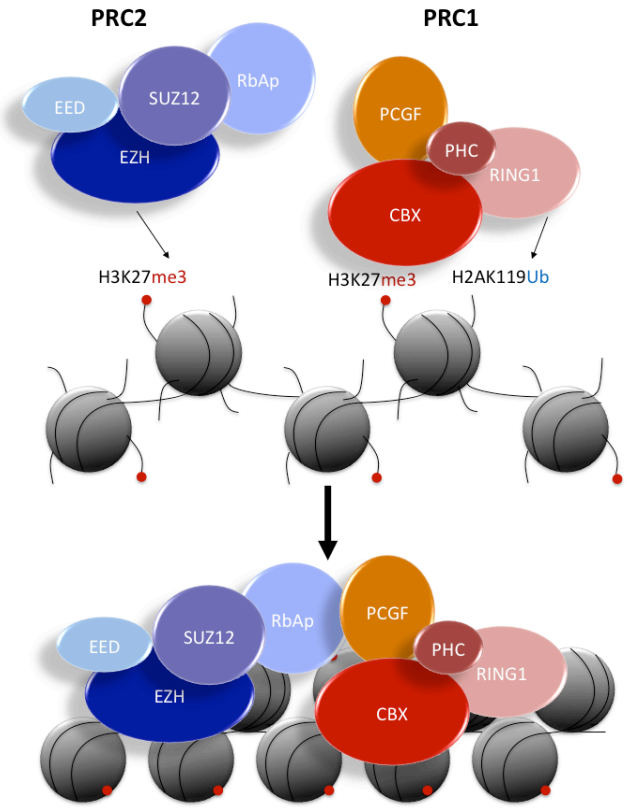

Figure Ch1- 3 Canonical Polycomb Complex.

The zinc finger Suz12, Eed, Ezh (SET domain with histone methyltransferase activity) and RbAp (histone binding domain) are the component of the PRC2 complex. PRC2 initiate the targeting of the sequence to be repressed and deposit H3K27me3. PRC1 composed of PCGF, PHC, Ring (ubiquitin ligase) and CBX (chromodomain that recognise H3K7me3) is required for the compaction of the sequence and stabilizing the silencing. However H2AK119Ub by Ring1 is not required for chromatin compaction (Eskeland et al., 2010b; Illingworth et al., 2015).

1.3.2.2 TADs and active hubs

The recent rise of chromosome conformation capture techniques allowed high throughput detection of interactions within chromosomes from the kilobase to the megabase scale. Those techniques are based on quantifying interaction between loci in a formaldehyde crosslinked cell. The crosslink allows to take a snapshot of the cell, then the genome is digested into smaller fragment followed by a ligation step that should favour proximity (in the 3D space) ligation. Next, 3D contacts can be quantified after an amplification step.

The first chromosome conformation capture (3C) experiment quantified one versus one interactions. The interaction from two loci at the time is monitored by PCR amplification (Dekker et al., 2002). Chromosome conformation capture-on-chip (4C) allows interactions detection from a bait sequence to a series of other sequences (one versus many) (Simonis et al., 2006). This also use PCR amplification and has been used to reveal specific enhancer-promoter interactions (Montavon et al., 2011). Chromosome Conformation Capture Carbon Copy (5C, many versus many) and Hi-C (all versus all) techniques are based on sequencing and quantifying reads. 5C allows to generate an interaction map of an entire region up to 1Mb (Dostie et al., 2006). The Hi-C technique allows genome-wide interaction mapping by combining the detection of sheared biotinylated 3C products by streptavidin beads with in depth sequencing (Lieberman-Aiden et al., 2009).

The use of Chromosome Conformation Capture techniques and its derivatives (3C, 4C, 5C, 3C-Seq, HiC) unveiled the genome's folding principles and exposed a repetitive structure of self-interacting domains (Bancaud et al., 2009; Lebedev et al., 2005; Lieberman-Aiden et al., 2009). Thereby, the mammalian genome appears to be confined into a succession of topologically associated domains (TADs). Those TADs subdivide the genome into ~1Mb domains in which chromatin interactions are enriched compared to inter-domain interactions (Figure Ch1-4A) (Dixon et al., 2012; Jin et al., 2013; Nora et al., 2012)

Recent work supports a model in which enhancers locate with their cognate promoters within the same topologically associated domains (TADs) and the ability of an enhancer to activate its target genes declines drastically beyond the TAD border (Anderson et al., 2014; Symmons et al., 2014). Furthermore, disrupting TAD boundaries has been shown to be able to re-wire enhancer and promoter communication leading to ectopic gene activation

(Lupiáñez et al., 2015) revealing that TADs functionally compartmentalise the genome to facilitate gene regulation.

This 1Mb size of TAD and maximum distance between enhancer and their cognate promoter might arise from the diffusion property of the chromatin itself. Indeed, as chromatin has a diffusion coefficient of ~1 x 10^{-4} $\mu m^2.s^{-1}$ in living human cells (Chubb et al., 2002), this defines a certain limit for enhancer-promoter interactions as it delimits a radius of constraint of ~ 0.5 µm. In case of a looping mechanism, enhancer and promoter sequences need to probe the space to find each other and for that to occur efficiently, the two sequences should be less than 0.5 µm apart which generally represents ~ 1 Mb. Therefore, chromatin motion might have provided a selective pressure for the maintenance of those 1Mb compartment.

Characterization of human chromatin using chromosome conformational capture techniques, strongly support the fractal globule structural model of the human genome (Lebedev et al., 2005; Mirny, 2011). That is to say that the three-dimensional chromatin architecture is organized as a repetitive folded entity from an active chromatin hub (Figure Ch1-4C) to a TAD (Figure Ch1-4A) to whole chromosome, that is subdivided in several TAD and that occupy its own territory (Figure Ch1-4B). However, recent imaging studies highlighted that chromosome folding diverge slightly from the ideal fractal-globule model as TADs size can vary and that TADs associates into higher compartments of the chromosome that are spatially arranged in a polarized fashion (Wang et al., 2016). Those last observations corroborate with FISH data that shows human p and q chromosome arms occupying non-overlapping, polarized chromosome territories (Boyle et al., 2011; Shopland et al., 2006).

TADs appear to be chromatin domains that define the functional landscapes of enhancers (Berlivet et al., 2013; Lupiáñez et al., 2015). Regardless of the mechanism of action, activation of gene expression mediated by an enhancer requires the establishment of protein-protein and protein-DNA complexes. As protein nuclear search depends on the laws of mass action, efficient gene activation via protein binding is sensitive to local concentration of both proteins and DNA. Therefore chromatin packaging is an important determinant of regulatory domains. Recent *in vivo* analysis of some transcription factors revealed that their prominent DNA binding mechanism, facilitated diffusion, seems to be a 3D diffusion alternated with 1D sliding along the chromatin fibre. A first model is to consider chromatin packaging as allowing the clustering of transcription factors binding sites and thus greatly reducing the

nuclear search space (Chen et al., 2014; Izeddin et al., 2014; Normanno et al., 2015; Woringer et al., 2014). Therefore multiple binding site for the same transcription factor on an enhancer will induce a local concentration of this same transcription factor. Constraining those proteins within the same chromatin hub would enable them to find the cognate promoter binding sites more easily and rapidly (Figure Ch1-4C). In this model, enhancer-promoter specificity relies on shared specificity and high affinity of their binding sites for the same transcription factors.

A second model uses a similar reasoning but with the inverse interpretation – suggesting that tight chromatin packaging reduces the nuclear space in which transcription factors are free to diffuse - excluded volume effect - (Woringer et al., 2014). Because of its viscoelastic properties, compact chromatin domains have been shown to affect chromatin motion and to significantly reduce the first encounter time for distant genomic regions (Lucas et al., 2014). This is consistent with the observation that pivotal developmental genes, such as *Shh*, and all their tissue-specific enhancers are located consistently within ~1 Mb chromatin domains that are physically very compact regardless of the tissue (Lettice et al., 2014; Williamson et al., 2012)**,** suggesting acquired physical mechanisms for TADs to prevent ectopic activation . We can imagine that depending on the cell type and the state of this locus, a chromatin hub might be compacted or not to either exclude potential activators binding or in contrary to increase the concentration of activators (Figure Ch1-4B).

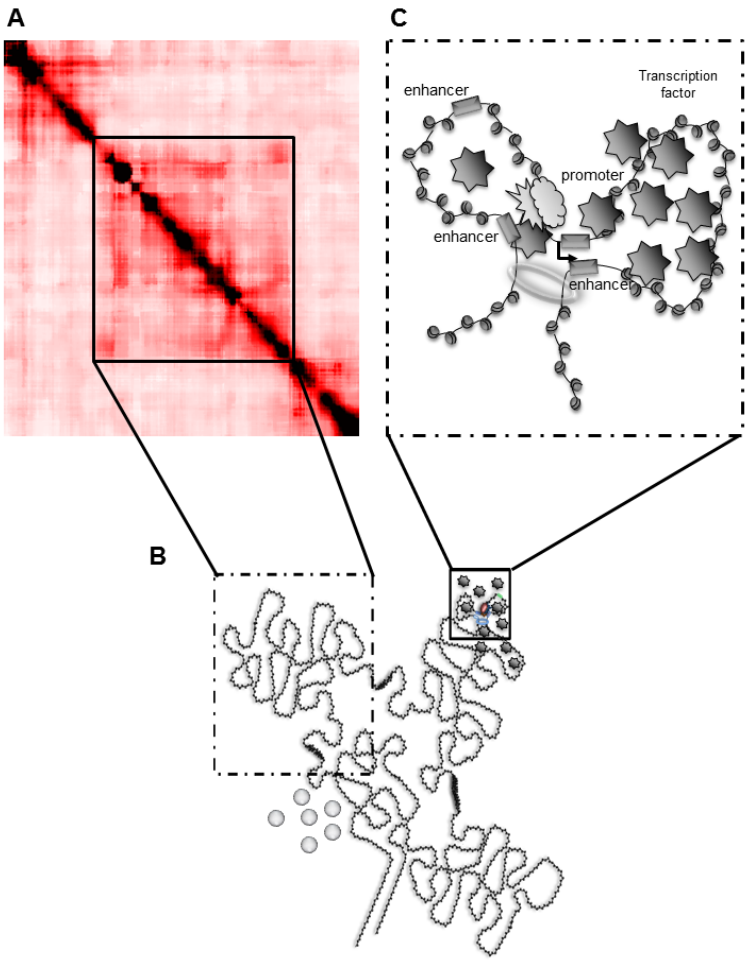

Figure Ch1- 4 TADs and Chromatin hubs.

A) Mouse embryonic stem cell (ESC) 5C heatmap revealing domain that self-interacts more, namely TAD. B) Schematic representation of more compact chromatin domains, such as inactive TADs or chromatin hubs, and their viscoelastic properties that can constrain chromatin motion and affect the dimensionality of transcription factor diffusion, and create either local concentration of transcription factors and activators or can physically exclude ectopic activators. C) Enhancers may just be constrained within the same nuclear domain (TADs or sub-TADs) as their target genes with recruiting protein and chromatin constrained diffusion then being sufficient to activate gene expression (Chromatin hubs).

In light of the previous section, how Polycomb interacting domains are related to the TAD segregation of chromatin remain unclear. Williamson et al. 2014 showed that the HoxD TAD is completely dependent on Polycomb but such mechanism is not true for every TAD. Surprisingly, most of the TADs are conserved between different cell lineages even when they contain genes that have different activation states and even when the underlying histone modifications change in parallel. In a general manner also, TADs structure remains unchanged when proteins responsible for H3K9 and H3K27 methylation are removed (Nora et al., 2012) but has been shown to change for the HoxD locus (Williamson et al., 2014).

Most of TADs studies emerged from FISH and "C" data that have low resolution (Dostie and Bickmore, 2012). Better sequencing quality and recent super-resolution imaging of TADs will perhaps in the future improve our current knowledge of those structures (Giorgetti et al., 2015).

Moreover, little is known regarding TADs molecular structures. CCCTC-binding factor (CTCF) and cohesin are well-described chromatin folding architects (Merkenschlager and Odom, 2013; Nichols and Corces, 2015; Ong and Corces, 2014; Parelho et al., 2008). More recently, TADs boundaries have been reported to be enriched for CTCF, cohesin and other architectural proteins (Dixon et al., 2012; Jin et al., 2013; Nora et al., 2012; Rao et al., 2014; Seitan et al., 2013; Sofueva et al., 2013; Zuin et al., 2014) with CTCF sites orientation dictating TADs architecture (Guo et al., 2015). Despite these advances the underlying molecular basis remains unclear. So far all that is known is that CTCF interacting sites are strongly shaping the position of TADs (and inner sub-TADs smaller self-interacting region contained in a TAD) which are consistent with the known characteristics of CTCF as an insulator of enhancer activity (Bell et al., 1999) and with promoting in other cases, long-range interaction and enhancing transcription (Xu et al., 2011).

1.3.2.3 Chromatin extrusion

At the whole chromosome scale, chromosome territory has been linked with changes in conformation and gene activity. Relocation of genes to the periphery or even the outside of visible chromosome territory often correlates with their activation status (Mahy et al., 2002; Volpi et al., 2000). FISH data and recent Hi-C interaction probing revealed that active and inactive domains generally segregate. Inactive domains tend to remain within the chromosome territories whereas active domains are less compact, locate to the chromosome territory periphery, and are more likely to form inter-chromosomal contacts with other active regions (Boyle et al., 2011; Lieberman-Aiden et al., 2009; Sexton et al., 2012; Yaffe and Tanay, 2011).

For example, Hox clusters reside in the internal part of their chromosome territories in embryonic stem cells (ESCs) and, upon activation, they loop out and reside outside of their chromosome territories. A similar event has been described for the *Shh* locus regarding the ZRS activity in the posterior side of the limb only (Amano et al., 2009) and with ectopic β-globin LCR (Noordermeer et al., 2008, 2011). However extrusion of the chromosome territory is not sufficient to induce gene activation (Morey et al., 2009). With for example the human α-globin genes that extend out of their chromosome territory irrespective of transcriptional status (Brown et al., 2006).

The functional role of looping out remain unclear as it does not induce activation *per se*, but as discussed in the previous section, this dynamic of chromosome territory extrusion and decondensation could allow more diffusion capacity. This could help to explore a larger nuclear space and decrease the first encounter time with transcription factors (Brown et al., 2008; Schoenfelder et al., 2010; Sutherland and Bickmore, 2009).

A parallel for chromosome extrusion phenomena could be drawn with heat-shock induced polytene chromosome puffs in Drosophila. Upon heat shock, Poly(ADP-ribose) polymerase-1 (PARP1) is strongly activated and parylates e.g. transfer of poly-ADP-ribose on very large regions in the polytene chromosome puffs which induce chromatin loosening (Tulin and Spradling, 2003; Tulin et al., 2002, 2003). PARP1 has been extensively linked with gene expression and regulatory regions (Nalabothula et al., 2015; Ogino et al., 2007; Petesch and Lis, 2012).

PARP1 is a chromatin remodeller that binds DNA via two zinc finger motifs and polymerizes long chains of poly (ADP-ribose) (PAR) from nicotinamide-adenine-dinucleotide (NAD+) to chromatin-associated acceptor proteins, core histones, histone variants, post-translationally modified histones, histone H1 or PARP1 itself (Gottschalk et al., 2009; Kim et al., 2004; Kraus and Lis, 2003; Petesch and Lis, 2008; Timinszky et al., 2009) (Figure Ch1-5).

The main histone acceptor of poly(ADP-ribose) is histone H1, which is then responsible for large relaxation of the chromatin architecture as it impairs H1's ability of compacting chromatin (Poirier et al., 1982). Histone H1 does not participate in the formation of the nucleosome but binds to the nucleosome and to the linker DNA and regulates the position of the nucleosome (Allan et al., 1981; Fan et al., 2005; Woodcock et al., 2006). H1 is the fundamental component of nucleosome positioning and stability and therefore an elementary architect of the chromatin structure via a very dynamic binding (Allan et al., 1980; Misteli et al., 2000; Thoma and Koller, 1977; Xiao et al., 2012). Therefore PARP1 is considered as an important chromatin architect as it leads to large scale chromatin loosening (Figure Ch1-5).

On the whole, chromatin architects encourage chromatin folding of domains following the same transcription regime. They compact together regions in the genome that are required to be silenced, and reciprocally enhance chromatin unfolding for domains that are to be transcribed.

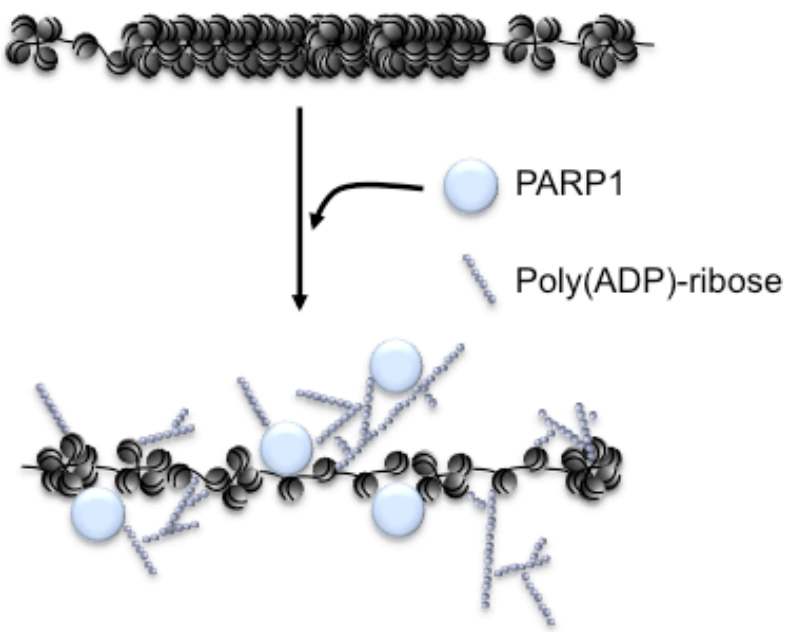

Figure Ch1- 5 Poly(ADP)-ribose Polymerase 1.

PARP1 transfers poly(ADP)-ribose onto the chromatin and chains of ADP ribose forming branching figures (Hayashi et al., 1983) leading to the loosening of the chromatin structure.

.

1.4 The interesting case of *Sonic hedgehog*

Tissue patterning is governed amongst other things by a subset of genes that encode signalling molecules called morphogens. Sonic hedgehog (*Shh*) as an important morphogen is tightly regulated by a group of enhancers. As shown in Figure Ch1-1, enhancers are positioned in a *Shh* regulatory domain that spans a large gene desert of 1Mb in the human genome, nearly 900 kb in the mouse and 1.4Mb in the opossum genome (Anderson and Hill, 2014; Epstein et al., 1999; Jeong, 2003, 2006; Lettice et al., 2002, 2003; Sagai et al., 2004, 2009; Yao et al., 2016).

SHH is a secreted signalling protein that imparts patterns of growth and identity to cells during many stages of embryonic development such as the developing central nervous system and the limb buds (Dessaud et al., 2008; Lettice et al., 2002, 2003; McMahon et al., 2003; Riddle et al., 1993).In limb buds, *Shh* is active in the zone of polarising activity (ZPA). From this posterior side of the limb to the anterior side, *Shh* is expressed as a gradient. If this gradient is disturbed it leads to ectopic expression of *Shh* in the anterior limb and will result in digit duplications which is the cause of congenital malformations such as preaxial polydactyly (PPD) (Riddle et al., 1993). Genetic analysis demonstrated that mutations in intron 5 of the *Lmbr1* gene lead to this ectopic *Shh* activation and thus to PPD (Lettice et al., 2002, 2003). This conserved region was named ZPA Regulatory Sequence (ZRS). Moreover, deletions and point mutations of the ZRS in mice, humans, cats and chickens all disrupted *Shh* and led to PPD or other limb defects (Furniss et al., 2008; Lettice et al., 2003, 2008; Maas et al., 2011; Sagai, 2005) confirming that the role of this enhancer is greatly conserved and establishing *Shh* as a text-book model for enhancer activity.

Shh is also very important in patterning the central nervous system (CNS) and brain. *Shh* is critical for a proper brain formation and also a Shh dorso-ventral gradient patterns the identity of motor neurons (Balaskas et al., 2012) (Figure Ch1-6A).

Regarding the brain, most of the known *Shh* neural enhancers have been identified through transgenic reporter assays in mouse. Shh-floor-plate enhancer-1 (SFPE1), -8kb from *Shh* transcription start site (TSS), SFPE2 and Shh-brain-enhancer-1 (SBE1) – both in the 2nd intron of *Shh*, show activity respectively in the ventral spinal cord (floorplate) and hindbrain or in the ventral midbrain, ventroposterior diencephalon and the Zona limitans intrathalamica

(ZLI) (Epstein et al., 1999; Jeong, 2003). Using BAC clones as reporter constructs, an enhancer trap assay that screened 1 Mb of the *Shh* regulatory region revealed Shh brain enhancer-2 (SBE2), SBE3 and SBE4 driving *Shh* expression in the ventral forebrain (ventral diencephalon for SBE2, ventral telencephalon for SBE4) (Jeong, 2006). More recently, using first pattern comparison from enhancer-reporter assay data base and motif recognition of similar enhancer sequence, a new SBE1-like ZLI enhancer has been annotated, SBE5 (Yao et al., 2016) (Figure Ch1-1).

Perturbation of *Shh* cis-regulation leads to severe defects in mammals. When the action of cis-regulatory regions is disrupted by translocations separating SBE2, 3 and/or 4 from *Shh*, it leads to a haplo-insufficiency phenotype similar to those resulting from autosomal dominant mutations in the coding region of *Shh* and cause diverse holoprosecencephaly –midline cleavage defect) (HPE) phenotypes (Figure Ch1-6B) (Belloni et al., 1996; Dubourg et al., 2007; Roessler et al., 1996). Additionally, a SBE2 point mutation results in loss of enhancer activity in the hypothalamus and causes HPE (Jeong et al., 2008). Together these cases highlight the importance of reporting and understanding new cis-regulatory elements, particularly in the context of understanding developmental disease.

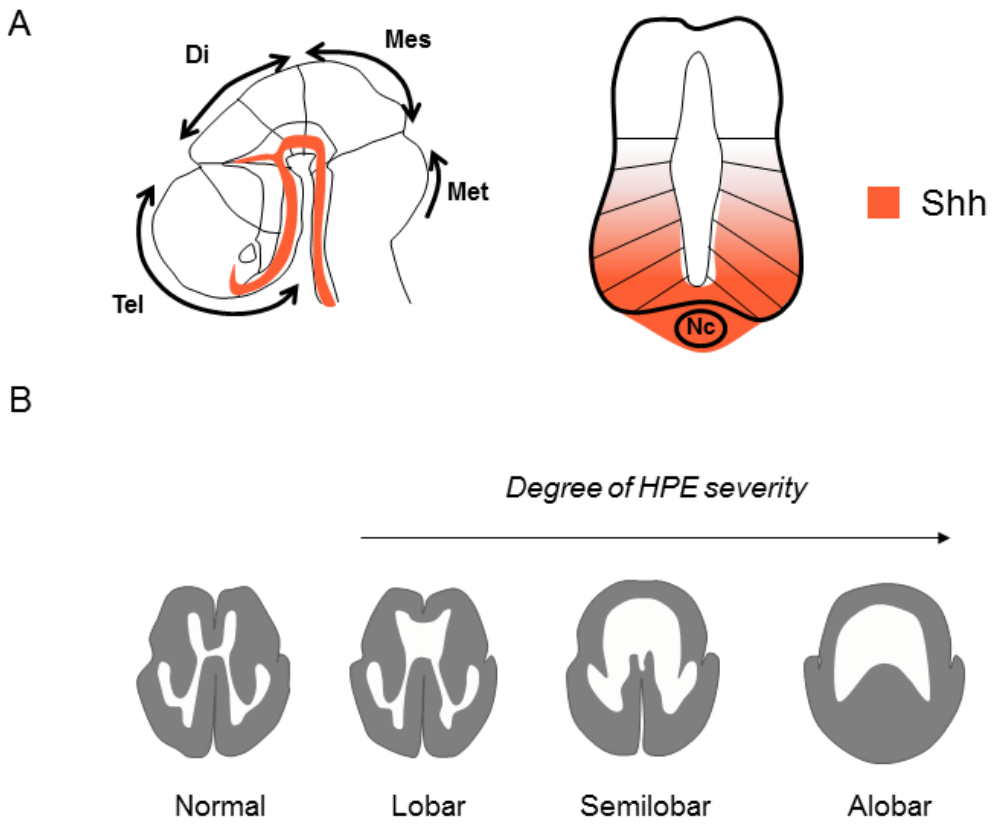

Figure Ch1- 6 Shh Signalling in a brain and gradient in the Vertebrate Neural Tube.

A) Left, *Shh* expression pattern in the developing mouse embryo and right, gradient from the notochord (Nc) and floor-plate of the spinal cord to the dorsal side of the neural tube leading to identity of motor neurons. B) Midline cleavage defect in various cases of holoprosencephaly (HPE).

1.5 PhD Aims

Most of the documentation about enhancer-promoter loops come from the strong detection of enhancer-promoter ligation products in the Chromosome Conformation Capture (3C) technique and its derivatives (4C, 5C, Hi-C etc.) (Montavon et al., 2011; Palstra et al., 2003; Tolhuis et al., 2002). Unfortunately, these "C" techniques are generated on a whole population and only reveal the average conformation and not the frequency of those interactions.

One of the first limitations of these techniques is the lack of information regarding a transient or stable conformation. Imaging technique such as fluorescence in situ hybridisation (FISH) allow the precise distribution of distances and thus interaction frequencies to be determined, as each allele is imaged separately. FISH has been used to study enhancer-promoter looping (Lettice et al., 2014; Williamson et al., 2012) and it could in some cases, when the frequencies observed are low, argue in favour of a more transient enhancer-promoter loop instead (Williamson et al., 2012).

Furthermore, recent side-by-side 5C and FISH comparison established that in some cases these techniques are not able to reproduce the same result, with physical proximity inferred from 5C heatmaps not observed by imaging (Williamson et al., 2014) and increased spatial co-localisation by FISH not reflected by increase of 3C signal detections (Williamson et al., 2016).

We could suppose that "C" techniques are powerful tools to study 3D folding of the chromatin as they reveal stable direct interactions. FISH could indicate in theory not only physical proximity by chromatin folding but also physical proximity due to condensation of the chromatin fibre into higher order structures (Figure Ch1-7). Therefore combining both techniques could allow having a better insight of the overall chromatin topology but also local chromatin fibre organisation.

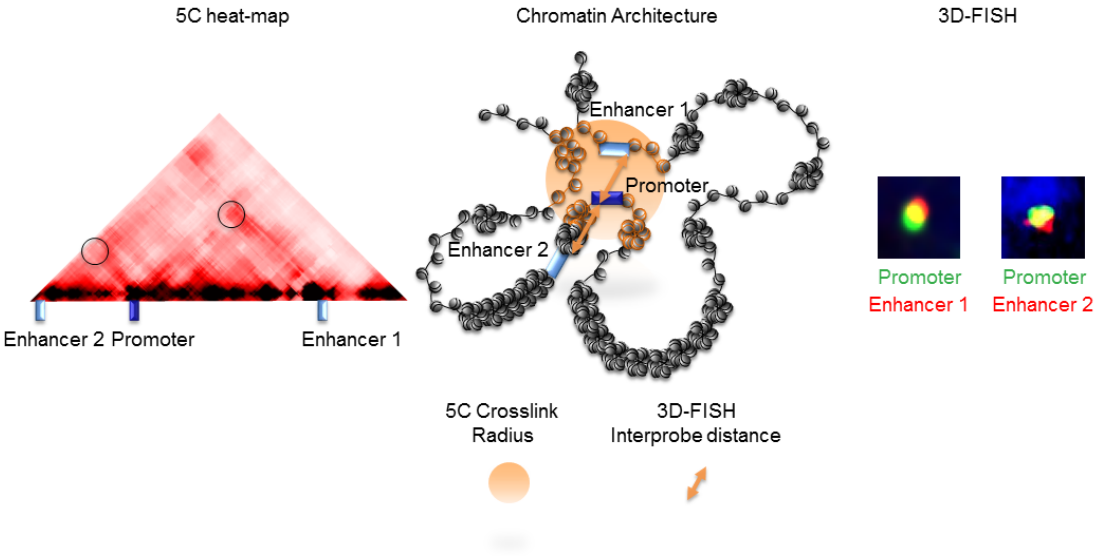

Figure Ch1- 7 5C heatmap and 3D-FISH contradictions.

Centre, schematic representation of a region with one promoter and two enhancers. The promoter is able to crosslink with Enhancer 1 due to a direct physical interaction that allows further detection of ligation products between those two sequences. But the promoter does not crosslink with Enhancer 2 (outside the orange circle representing 5C crosslink radius) as the two sequences do not directly interact due to higher chromatin fibre organisation. However, even though both these sequences are located far away from the promoter in genomic terms, they both show a strong co-localisation signal by FISH as the 3D-FISH interprobe distance is the same (orange arrow). Left, 5C heatmap illustrating differential interaction between Enhancer 1, promoter and Enhancer 2. Right, 3D-FISH representing same distances for each couple of probes.

Combining both tools, studies of Shh-ZRS communication showed that ZRS activate *Shh* following the looping model (Amano et al., 2009; Williamson et al., 2016).

Shh is an important Polycomb target and its conformation in non-expressing cells such as ESCs is very compact. The Shh regulatory domain is contained in a single TAD with Shh located at one border and the series of enhancers located upstream its TSS in the TAD. Shh TAD comprises several CTCF sites that interact together and shape inner-TAD interactions (Williamson, Lettice et al., *unpublished*). Shh-ZRS has been well studied, however, how Shh brain enhancers activate *Shh* remain unknown.

The aim of my PhD is to dissect how *Shh* communicates with its active neural enhancer during differentiation. SBE2 has a known effect on cases of holoprosenchapaly linked with missexpression of Shh, however, how SBE2 interacts with Shh is unknown. Understanding the mechansim by which SBE2 activates Shh will help us to conceive why a single point mutation in one of its transcription factor binding site (SIX3) leads to major deregulation of *Shh* expression.

To answer the aim, I used a differentiation system from embryonic stem cells (ESCs) where *Shh* is not active to neural progenitor cells (NPCs) where *Shh* is activated to follow enhancer-promoter communication using microscopy and molecular biology techniques. In Chapter 2, I annotated a new distal *Shh* enhancer active in NPCs named Shh-brain-enhancer-6 (SBE6). Chapter 3 describes how I first observed a surprising interaction between *Shh* and its brain enhancer. Chapter 4 describes the dissection and analysis of the observation made in Chapter 3 using synthetic techniques. Chapter 5 presents the mechanism(s) involved in this unique interaction. Finally Chapter 6 will be the discussion of all my work and conclusion of this thesis.

Chapter 2: Shh-Brain-Enhancer 6 (SBE6), a novel long-range enhancer driving Shh expression in neural progenitor cells

2.1 Chromatin signatures of enhancers

In order to expose crucial transcription factor binding sites (TFBSs), enhancers and promoters share properties of chromatin structure including increased chromatin accessibility, the presence of chromatin modifications associated with open chromatin and the transcription of short non-coding RNAs.

Traditionally DNAse-I hypersensitive site (DHS) mapping has been used to identify potential active enhancers by measuring the accessibility of DNA to nuclease digestion (Gross and Garrard, 1988). More recently, accessibility of DNA to Tn5 transposition forms the basis of ATAQ-seq (Buenrostro et al., 2013) which gives similar maps of chromatin accessibility following a more straightforward experimental procedure. Together these approaches identify potential regulatory elements *in vivo*.

Complementing chromatin accessibility data, a host of posttranslational modifications of histone proteins have been correlated with enhancer activity (Heintzman et al., 2009; Pradeepa et al., 2016; Roh et al., 2007, 2005; Taylor et al., 2013). Developmental enhancers are pre-marked by H3K4me1 and show a gain of H3K27ac when they become active (Calo and Wysocka, 2013). The H3 acetyltransferase p300 and other HATs also binds at enhancers, as well as promoters (Heintzman et al., 2007, 2009; Krebs et al., 2011). The correlation between the histone modifications and enhancer activity is sufficiently close that these marks are now used to identify potential enhancers in the genome (Bonn et al., 2012), independent of chromatin accessibility data.

Even though new genetic approaches can clearly identify bone fide active enhancers in a genomic region (Bonn et al., 2012), they are unlikely to be capable of identifying all such elements comprehensively (Pradeepa et al., 2016; Taylor et al., 2013). Approaches for functionally describing new regulatory elements remain time-consuming, as enhancer reporter assay has to be done in a case-by case manner. However, molecular approaches including chromatin accessibility, histone modifications and transcription can give a strong indication for predicting enhancer activity *in vivo* which can then be tested on a case-by-case basis.

In order to precisely study *Shh* and its brain-enhancer, I chose to establish which neural enhancers for *Shh* are activated in a traceable *ex vivo* differentiation system. I therefore used a mouse embryonic stem cell (ESC) line - 46c that contains a GFP reporter knocked in to the *Sox1* locus (Abranches et al., 2009; Pollard et al., 2006; Ying et al., 2003), that allows the purification by FACS of Sox1-GFP^{+ve} neuroepithelial progenitor cells (NPCs). Sox1-GFP^{+ve} positive NPC appear after day 3 of N2B27 differentiation, and from day 3 to 7, expression of *Shh* and *Nestin* increase whilst *Oct4* mRNA levels progressively decrease (Figure Ch2 -1).

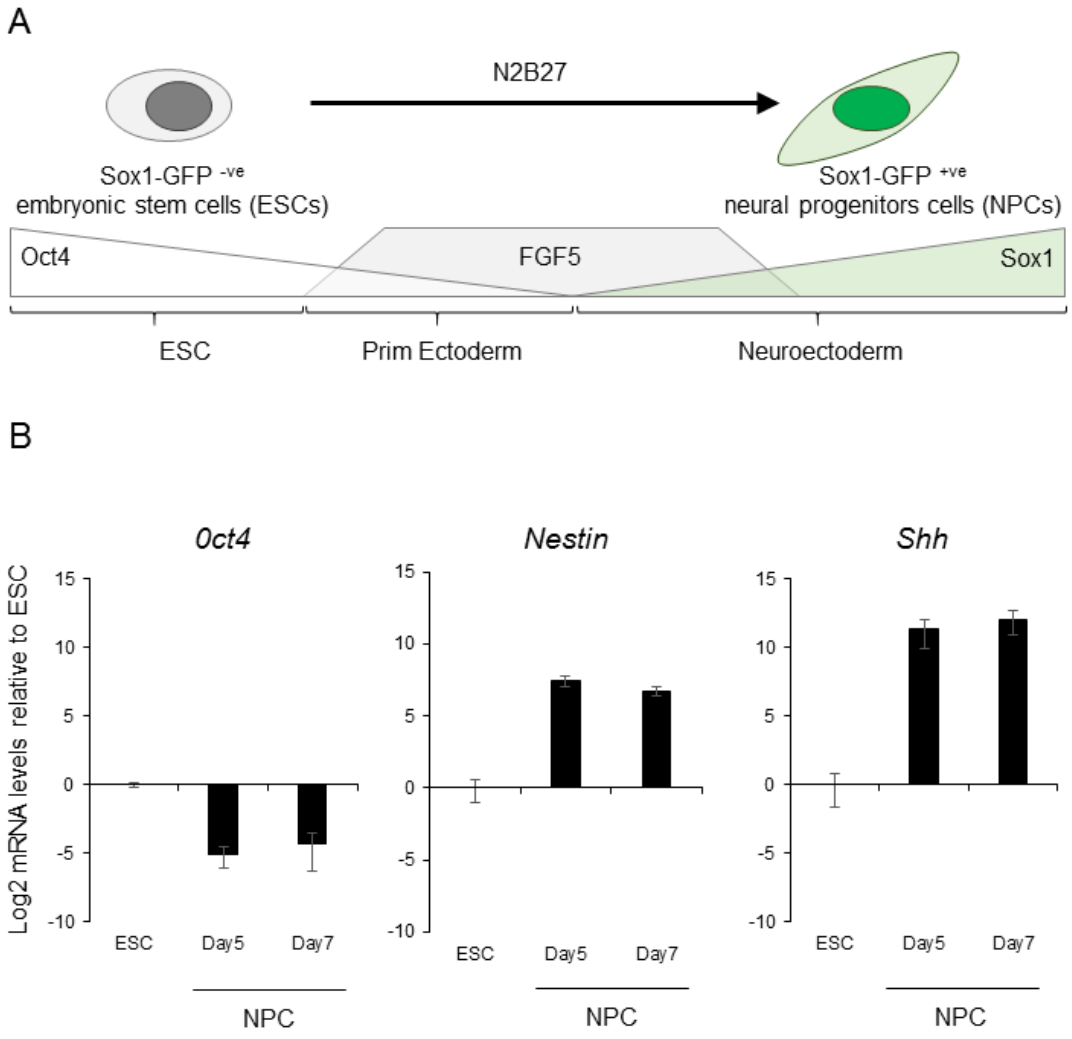

Figure Ch2- 1 N2B27 Neural Differentiation.

A) Schematic showing N2B27 differentiation of Sox1-GFP^{-ve} 46c mouse embryonic stem cells (ESC) into; first primitive ectoderm as Oct4 levels decrease and Fibroblast growth factor 5 (FGF5) levels rise, and then further into neuroectoderm as FGF5 levels start to decrease and Sox1-GFP^{+ve} neural progenitor cells (NPC) increase. B) qRT-PCR showing mean ± standard error of the mean (s.e.m.) levels of log2 *Oct4*, *Nestin* and *Shh* mRNA levels in ESC and NPCs after 5 and 7 days of differentiation. Expression levels are relative to *Gapdh* and normalised to one biological replicate of data from ESCs. Data are from 2 biological replicates and technical triplicates.

2.2 Analysis of histone modification changes in the *Shh* regulatory region during neural differentiation

Using native chromatin immunoprecipitation (ChIP) on chip using genomic microarrays that tile the whole regulatory region of *Shh,* Betül Hemikoglu-Balkan, a post-doc in the lab, assayed histone modifications H3K4me1 and H3K27ac in ESCs where *Shh* is not expressed and in day 5 Sox1-GFP^{+ve} NPCs were *Shh* is expressed. The ChIP-on-chip microarrays have been processed by Graeme Grimes (IGMM Bioinformatics). Surprisingly, we found little change in the assayed histone modifications over the well-known Shh-brain-enhancers (SBE2, 3 and 4) (Figure Ch2-2A). Our attention got caught particularly by a small highly conserved region located around a 100 kb upstream of the *Shh* TSS that has no evidence of active enhancer marks in ESC but gains both H3K4me1 and H3K27ac marks upon neural differentiation (Figure Ch2-2).

In silico analysis of this region performed by Philippe Gautier - IGMM Bioinformatics core - allowed us to identify two smaller regions (~1kb) under this ChIP peak, corresponding broadly to regions with higher conservation in Vertebrates and visualized in the "Vertebrate Multiz Alignment & Conservation/Multiz Alignments and Conserved elements" tracks in the UCSC genome browser. This resulted in two "core regions" that are strongly conserved in mammals and birds – though not fish -. I named these putative NPC enhancer SBE6.1 (mm9 co-ordinates Chr5: 28889688-28890461, 96048bp upstream of Shh TSS) and SBE6.2. Chr5: 28893935-28895000, 100295bp away from Shh) (Figure Ch2-2B).

Searching for transcription factor binding sites (TFBS) on those two sequences using rVISTA algorithm, Jaspar and RSAT scans to align the mouse and human orthologous sequences, with the default sequence aligner (LAGAN) and default parameters, allowed us to confirm the presence of binding sites of known forebrain transcription factors such as ETV1, SP8, VAX1 and DLX6 (Nord et al., 2015) (Table 1).

Interestingly SBE6.1 sequence is entirely included in a recently described lung and gut regulatory element for *Shh* expression in mouse embryos called SLGE (chr5: 28889230-

28890979) (Tsukiji et al., 2014), raising a question of multiple regulatory activity in this enhancer. SBE6.2 has not previously been identified or studied.

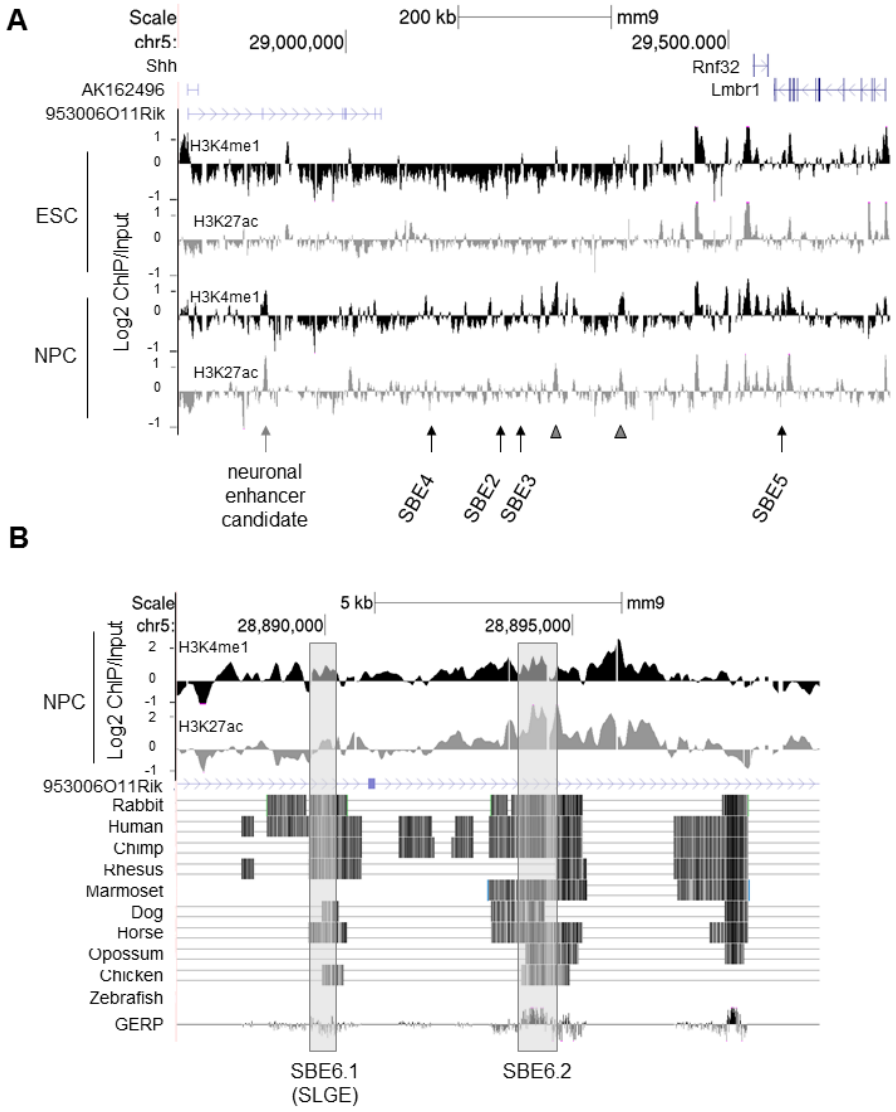

Figure Ch2- 2 Chip-on-chip for developmental enhancer marks on the Shh regulatory region.

A) Mouse chromosome 5 mm9 (chr5:28,782,000-29,702,000) view of log2 H3K4me1 and H3K27ac ChIP on ESC and day 5 NPC relative to input revealing a region that displays huge gain of enhancer marks in NPC indicated as neural enhancer candidate with grey arrow. Black arrows indicate well-known Shh brain enhancers and arrowheads point at regions undergoing gain that are not further investigated here. B) Zoom on the region (chr5:28,885,000-28,900,000) with conservation tracks displaying two smaller core regions named SBE6.1 spanning partially Shh lung and gut enhancer (SLGE) and SBE6.2 highlighted in grey.

Predicted putative sites in mouse SBE6.1

Model ID	Model name	Score	Relative score	predicted site sequence
MA0761.1	ETV1	5.477	0.81918712	AGAGGAAGGG
MA0761.1	ETV1	6.665	0.84161589	GAAGGAAGTA
MA0722.1	VAX1	6.22	0.85945793	CTCATGAA
MA0722.1	VAX1	5.993	0.8530141	TTCATGAG
MA0882.1	DLX6	4.104	0.82019633	TCAATGAG
MA0722.1	VAX1	5.676	0.84401544	CTCATTGA
MA0722.1	VAX1	6.194	0.85871987	TCAATGAG
MA0722.1	VAX1	5.696	0.84458318	CTAAAGAC
MA0761.1	ETV1	5.067	0.81144656	GAAGGATGTG
MA0747.1	SP8	8.565	0.84123614	CACAACCCCACT
MA0747.1	SP8	8.342	0.83709184	CCCTCCCCCACC
MA0761.1	ETV1	4.847	0.80729308	AAGGGAAGTT
MA0747.1	SP8	6.623	0.80514545	AGCACTCCCCTC
MA0882.1	DLX6	4.159	0.82162245	CCCATTTC
MA0761.1	ETV1	5.94	0.8279283	TCAGGAAGAG
MA0070.1	PBX1	9.456	0.82183775	AGATCACTCAAG
MA0722.1	VAX1	4.258	0.80376278	CACATGAC

Predicted putative sites in mouse SBE6.2

Model ID	Model name	Score	Relative score	predicted site sequence
MA0882.1	DLX6	4.778	0.83767285	ACAATAAA
MA0722.1	VAX1	4.603	0.81355627	TTTATTGT
MA0747.1	SP8	7.915	0.82915635	TCCACCCCTACA
MA0722.1	VAX1	4.873	0.82122074	CAAATGGC
MA0722.1	VAX1	4.775	0.81843882	TTCATTTC
MA0882.1	DLX6	3.662	0.80873546	CATATTAT
MA0722.1	VAX1	4.144	0.80052667	TAAATCAC
MA0747.1	SP8	9.031	0.84989642	GTCACACCCCCA
MA0722.1	VAX1	5.718	0.84520769	CTTATGAC

Table 1 Jaspar scores for forebrain TFBS in SBE6.1 and SBE6.2 in mouse.

Two other sequences beyond SBE3 also show a gain of active enhancer marks (arrowheads in Figure Ch2-2A). To look for evidence that these sequences are potential regulatory elements that are active in the brain, I examined regulatory segmentation built from ChromHMM or Segway using ENCODE data from various tissues such as E14.5 mouse brain. I could distinctly visualize known and unknown enhancers at the Shh regulatory region (Figure Ch2-3). This first confirmed the activity of SBE2, SBE3 and SBE4 in the mouse brain, which is not striking in the H3K27ac ChIP on NPCs (Figure Ch2-2). This also confirmed the overlapping of SLGE and SBE6.1 as these sequences seem to be active in epithelial linings but also in the brain (Figures Ch2-3 & Ch2-4).

But the most astonishing observation is that the Shh regulatory region muss possess hundreds of yet non-annotated brain enhancer as looking at a better resolution, even peak #2 - but not peak #1 - from Figure Ch2-2 and Ch2-3 seems to be a putative brain enhancer (Figure Ch2-4).

Using this regulatory segmentation on Ensembl browser I could see that both peak#1 and peak#2 are predicted to be much shorter brain enhancer, therefore I focussed on SBE6.1 and SBE6.2 to carry comprehensive characterisation of their enhancer activity.

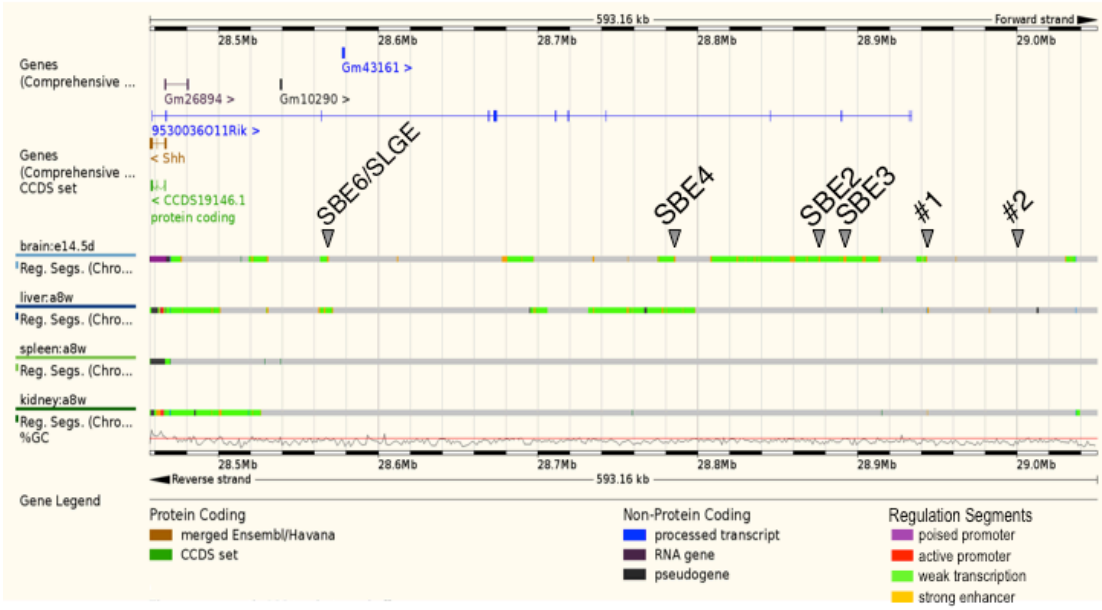

Figure Ch2- 3 Chromatin state discovery and characterization (ChromHMM) in mouse and human Shh regulatory region.

Ensembl *mus musculus* version 84.38 (GRCm38.p4) view of chr5:28456840-29050000 with regulatory feature tracks from ChromHMM. Grey arrowheads indicate previously described enhancers and two putative sequences indicated in Figure Ch2-2. Colour code from ChromHMM with purple for poised promoter, red for active promoter, bright green for weak transcription and orange for strong enhancers.

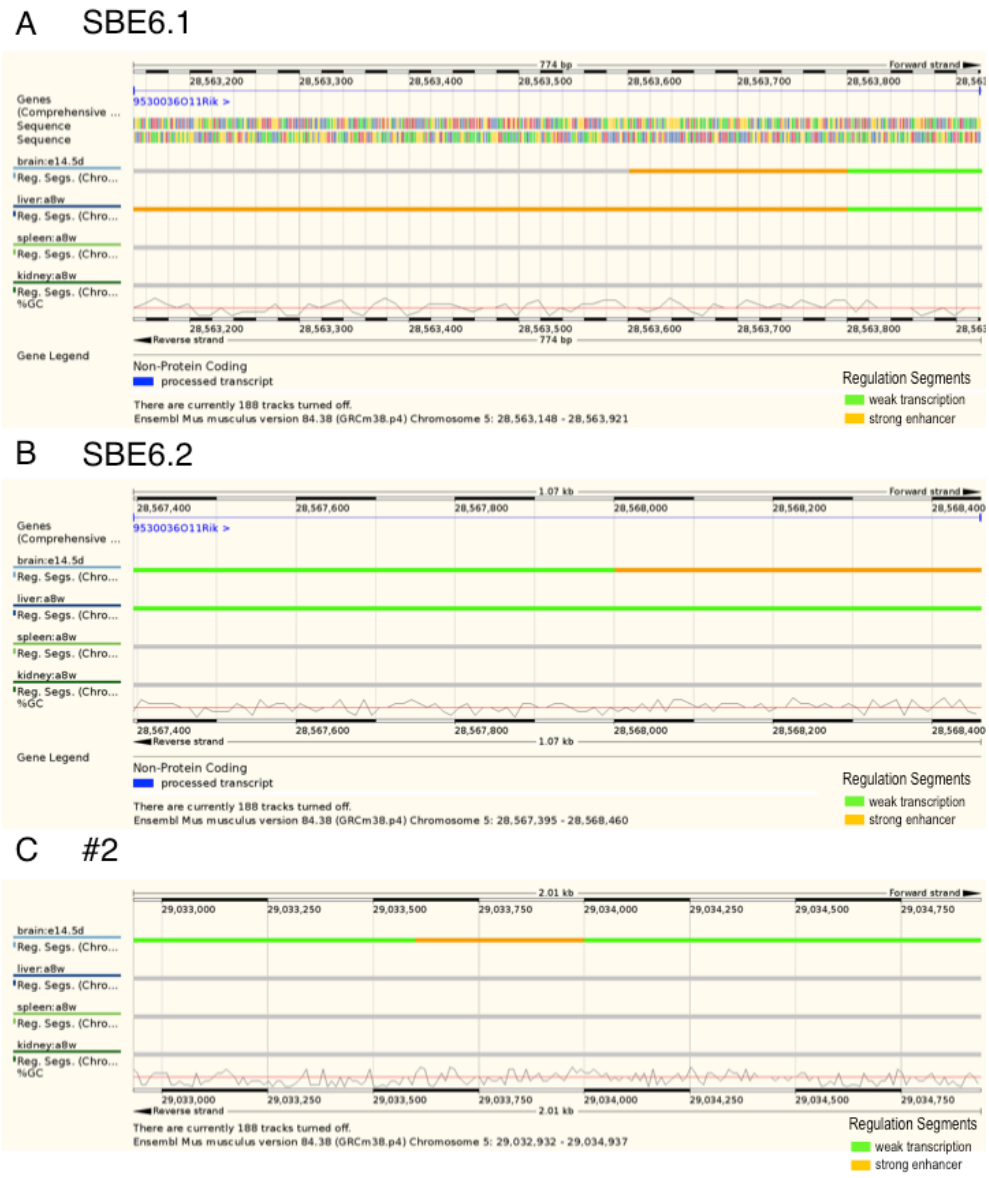

Figure Ch2- 4 Chromatin state discovery and characterization (ChromHMM) in SBE6.1, SBE6.2 and peak #2.

A) Ensembl *mus musculus* version 84.38 (GRCm38.p4) view of SBE6.1 with regulatory feature tracks from ChromHMM B) Ensembl *mus musculus* version 84.38 (GRCm38.p4) view of SBE6.2 with regulatory feature tracks from ChromHMM C) Ensembl *mus musculus* version 84.38 (GRCm38.p4) view of putative peak #2 with regulatory feature tracks from ChromHMM.

2.3 Exploring the enhancer activity of SBE6.1 and SBE6.2

2.3.1 Zebrafish reporter assay

To test the putative enhancer activity of SBE6.1 and SBE6.2, I used a zebrafish Tol2 transposon based enhancer-reporter transgenic assay in which the test element is juxtaposed to a minimal promoter driving the expression of a reporter gene, which encodes for eGFP or mCherry. This reporter assay in zebrafish has been shown be able to detect the expected expression pattern for wild-type and mutated enhancers (Bhatia et al., 2015; Jeong et al., 2008).

Shipra Bhatia, a post-doc in our lab helped me designing the primers (listed in Chapter 7) to amplify and clone the putative enhancer in plasmids containing either GFP or mCherry reporter gene. Together we cloned four constructs; SBE6.1-GFP, SBE6.2-mCherry, SBE6.1 negative control-mCherry and SBE6.2 negative control-GFP. The sequences used as negative controls are located the same distance away from Shh TSS as SBE6.1 and SBE6.2 (96048bp and 100295bp downstream of Shh TSS, in the opposite direction). We also cloned SBE4-GFP in order to visualize a positive brain enhancer but the data has not been acquired for this condition. Together with Shipra Bhatia, I built the different constructs, injected transient zebrafish embryos and Shipra Bhatia then bred the F0 transgenic animals to make stable transgenic lines. Together, we performed the confocal imaging of the transient and stable transgenic embryos.

Using this assay, SBE6.1 enhancer activity was detected mainly in the ventral part of the developing forebrain of the zebrafish embryos in four independent transgenic lines, with some punctual signal I dorsal part. However SBE6.2 failed to drive a consistent reporter gene expression with forebrain specific activity noted in only one out of the four independent transgenic lines generated (Figure Ch2-5 and Table 2).

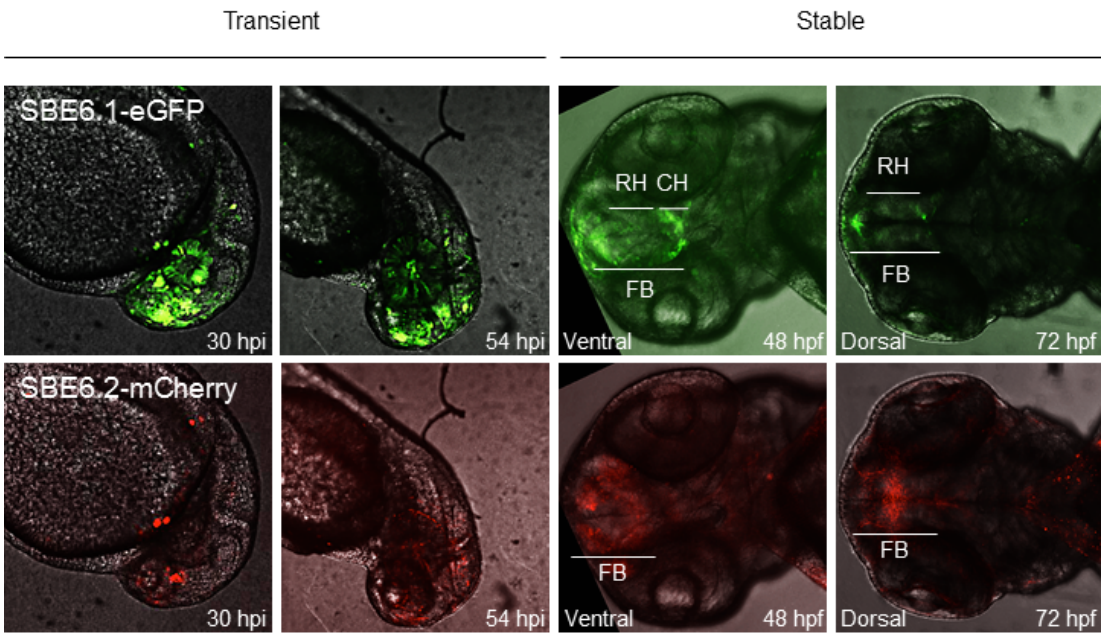

Figure Ch2- 5 SBE6.1 and SBE6.2 enhancer reporter assay during zebrafish development.

Confocal imaging of zebrafish embryos displaying Tol2 reporter assay with GFP and mCherry revealing respectively where SBE6.1 and SBE6.2 drive expression in transient injections after 30 or 54 hours post injection (hpi) (two left panels) and in stable transgenic lines 48 and 72 hours post fertilization (hpf) (two right panels). SBE5.1 and SBE5.2 show enhancer activity in the zebrafish forebrain (FB). Reporter gene expression is detected in the rostral hypothalamus (RH) and caudal hypothalamus (CH) of the forebrain (FB) in ventral and dorsal views.

Transgene *	Reporter used in transgenic assay #	Stable transgenic lines analysed	Sites of reporter expression driven by the element	Tissue-specific activity of the CRE observed in 100% of transgenic lines analysed
Shh-SBE6.1	eGFP	4	Forebrain (4/4; 100%) Pectoral fin (1/4; 25%) Heart (1/4; 25%) Retina (1/4; 25%)	Forebrain
Shh-SBE6.2	mCherry	4	Forebrain (1/4; 25%) Otic vesicle (1/4; 25%) Olfactory placode (1/4; 25%) Neural tube (1/4; 25%)	None
Shh-SBE6.1-negative ctrl	mCherry	2	Faint ubiquitous signal	None
Shh-SBE6.2-negative ctrl	eGFP	3	Notocord (1/3; 33%) Heart (1/3; 33%) Ubiquitous (1/3; 33%)	None

* Species of origin: *Mus musculus*

Model organism for reporter transgenic assay: *Danio rerio*

Table 2 Cis-Regulatory Element (CRE) driven transgene expression sites in F1 zebrafish embryos obtained from multiple independent stable transgenic F0 lines

2.3.2 Mouse LacZ reporter assay

As only SBE6.1 showed a consistent activity in the developing brain of the vertebrate embryo, I pursued my investigations using only SBE6.1. With the help of Shipra Bhatia I constructed a SBE6.1-LacZ construct in order to have a clearer indication of SBE6.1 enhancer activity in the developing mouse brain. The cloning of genomic regions containing the putative enhancer, generation of enhancer-reporter constructs and generation and further analysis of transgenic lines was carried out as previously described by Shipra and I (Ravi et al., 2013). Shipra Bhatia took care of establishing two stable lines with the help of David Read from Ian Adams's team. I performed the LacZ staining of the embryos. Unfortunately of those two lines; only one consistently expressed LacZ and from that unique line I managed to get only one embryo displaying LacZ staining (Figure Ch2-6).

With the help of Laura Lettice and Paul Devenney in Robert Hill's team, we also performed transient injection of the SBE6.1-LacZ construct into mouse embryos (Figure Ch2-6). The team of the CBS-IGMM transgenic Facility performed injections for both stable and transient assays. Laura and Paul revealed the LacZ staining from transient injection at stage E11.5. However, here again the construct gave a very low number of positive transgenic animals. Unfortunately from transient injection we also managed to visualize a proper LacZ signal in only one embryo. Subsequent analysis of that SBE6.1 in mouse LacZ reporter assay also revealed the enhancer activity in the developing brain of the mouse embryo but in that case expression is recorded in the mesencephalon (midbrain) of the mouse (Figure Ch2-6). SBE6.1 also showed activity in the floor plate of the spinal cord of the mouse embryo as well as in superficial diencephalon cells (Figure Ch2-6).

It is hard to conclude anything concrete from only two mouse transgenics. However the strong similarities between those two embryos could reassure us that SBE6.1 enhancer is active in the developing vertebrate brain with a variation from a forebrain pattern in the zebrafish where it is not conserved, to a floor plate and ventral midrain expression in mouse embryos.

Figure Ch2- 6 Ch2-6 SBE6.1-LacZ reporter assay in mouse developing embryo stage E11.5.

Top, external view of the LacZ staining SBE6.1 stable (left) and transient transgenics (right). Arrowhead pointing at signal in the superficial forebrain (FB), near the limit with the midbrain (MB). Bottom left, Shh in situ in mouse embryo E11.5 displaying Shh expression in the forebrain (telencephalon, diencephalon), (FB), midbrain (caudal diencephalon and zona limitans intrathalamica) (MD) and (ZLI), and hindbrain (HB). Middle, sagital section of the transient embryo with arrowhead pointing at SBE6.1-LacZ staining in the mouse midbrain and left, transversal section of the spinal cord with SBE6.1-LacZ signal in the floor-plate.

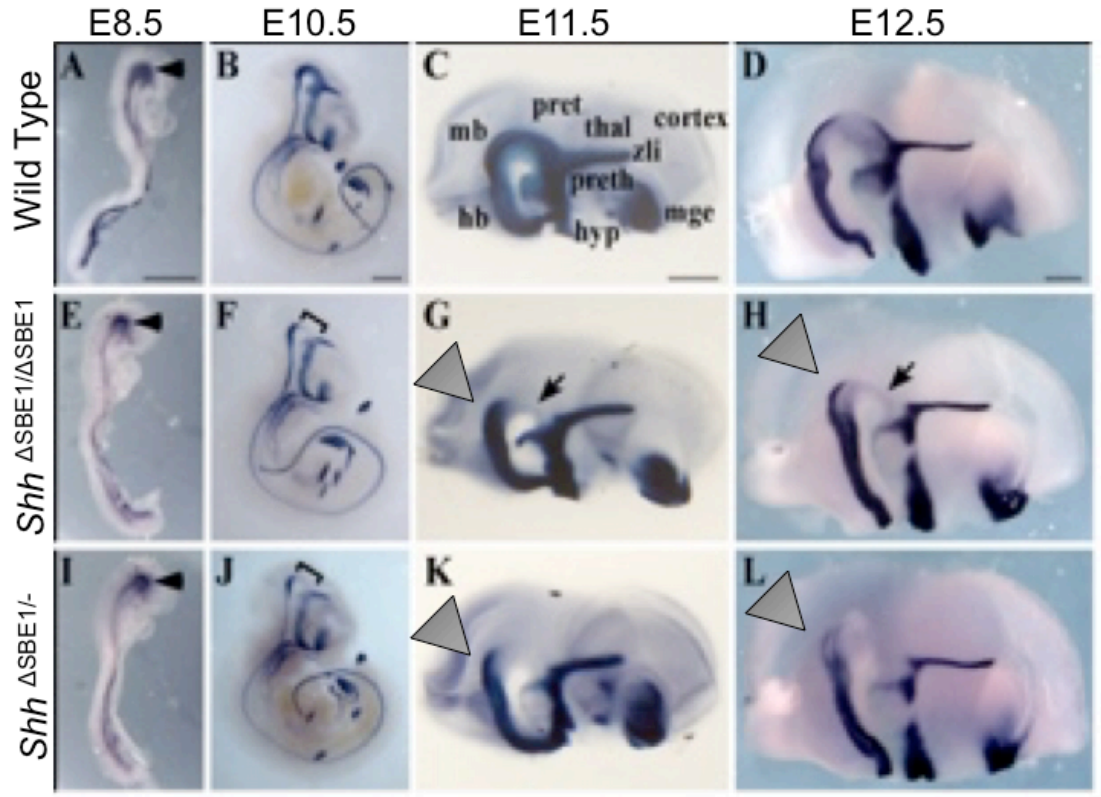

Figure Ch2- 7 (An) other enhancer(s) driving Shh expression in the midbrain.

Adapted from Figure 1 Maintenance of *Shh* expression is dependent on SBE1 of (Jeong et al., 2011). Grey arrowheads point at faint residual Shh expression in the midbrain in wild-type embryos or after SBE1 deletion. *Shh in situ* hybridization in wild type, Shh $^{\Delta SBE1/\Delta SBE1,}$ and Shh $^{\Delta SBE1/-}$ mouse embryos at various stages of development (E8.5, E10.5, E11.5 and E12.5). hb, hindbrain; hyp, hypothalamus; mge, medial ganglionic eminence; mb, midbrain; pret, pretectum; preth, prethalamus; thal, thalamus; zli, zona limitans intrathalamica.

The different results from SBE6.1 transgenic assays in zebrafish and mouse, forebrain vs. midbrain is difficult to explain. From a conservation point of view the mouse result should be more representative of SBE6 activity as it is not at all conserved in zebrafish. Thus the difference of activity pattern could be due to the fact that zebrafish proteins might recognize SBE6.1 as forebrain like transcription factor binding site in opposition to the mouse where the activation will be more accurate.

Comparing my data on SBE6.1 with published transgenic data for SBE1 (Jeong et al., 2011 Figure 1), another Shh midbrain enhancer active in caudal mesencephalon and currently thought to be the unique enhancer driving *Shh* expression in those tissues, it seems that the complete *Shh* midbrain expression is driven by yet another unknown Shh brain enhancers as, after SBE1 deletion, a faint residual *Shh* expression remains in the midbrain (Figure Ch2-7) (Jeong et al., 2011).

This could further indicate SBE6.1 activity in the midbrain, as the pattern is highly similar, or the other midbrain enhancer could reside somewhere else in the Shh regulatory region, as Figure Ch2-3 identifies many other non-annotated enhancer are active in the mouse brain.

2.4 Role of SBE6.1 and SBE6.2 during neural cell differentiation

In order to determine the role of SBE6.1 and SBE6.2 in their native genomic context, I used CRISPR/Cas9 to generate homozygous SBE6.1-/- and SBE6.2 -/- cell lines the 46c ESC genome (Figure Ch2-8). The guide RNAs were designed to delete the entire SBE6.1 or SBE6.2 sequence highlighted in grey in Figure Ch2-8. The PAM sequences were oriented divergently from the sequence to be deleted in order for the wild-type Cas9 from the pX458 plasmid to efficiently generates double strand breaks and chew DNA on the outside of the sequence.

I was able to generate two independent homozygote SBE6.1 -/- cell lines as well as three independent homozygote SBE6.2 -/- cell lines. For each condition, the independently generated clones displayed a reduction in the size of a PCR fragment amplified across the region, consistent with a homozygote deletion (Figure Ch2-8). The homozygote deletions have been further confirmed by sequencing.

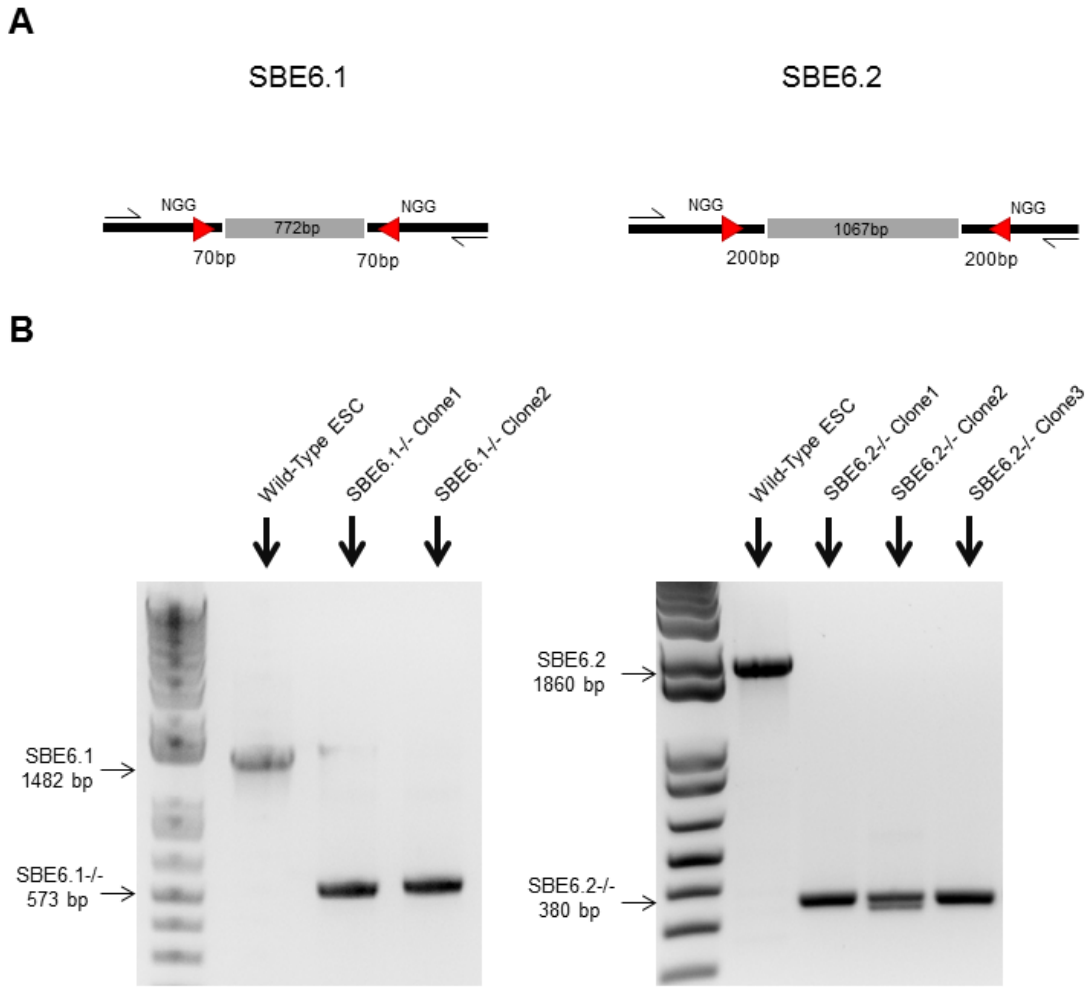

Figure Ch2- 8 Genomic PCR of SBE6.1-/- and SBE6.2 -/- ESC lines.

A) Illustration showing the SBE6.1 and SBE6.2 sequences with the gRNA targeting for deletions using CRISPR/Cas9 and primers used for genotyping. B) Left, Gel showing PCR amplification of SBE6.1 sequence in SBE6.1-/- ESC clone 1 and 2. Right, Gel showing PCR amplification of SBE6.2 sequence in SBE6.2-/- ESC clone 1, 2 and 3, independently generated displaying. Band size measured with 1kb+ DNA ladder (Invitrogen).

By deleting SBE6.1 and SBE6.2 in ESC, I could study the direct effect of these deletions on *Shh* expression and delineate if those enhancers are specific to *Shh* activation in NPCs. Neural differentiation into Sox1-GFP^{+ve} NPCs was not disturbed in either deletion cell line, as judged by the proportion of Sox1-GFP^{+ve} NPCs recovered after FACS and quantitative RT-PCR for *Oct4* and *Nestin* mRNAs (Figure Ch2-9A). NPCs differentiated from the SBE6.1-/- deleted cell lines failed to activate *Shh* expression to a similar level to wild type cells (Figure Ch2-9B). In contrast, average *Shh* mRNA levels in NPC differentiated from SBE6.2 -/- ESCs were marginally reduced relative to wild type but this was not statistically significant (Figure Ch2-9B).

I therefore concluded that SBE6.1 directly enhances *Shh* expression in NPC.

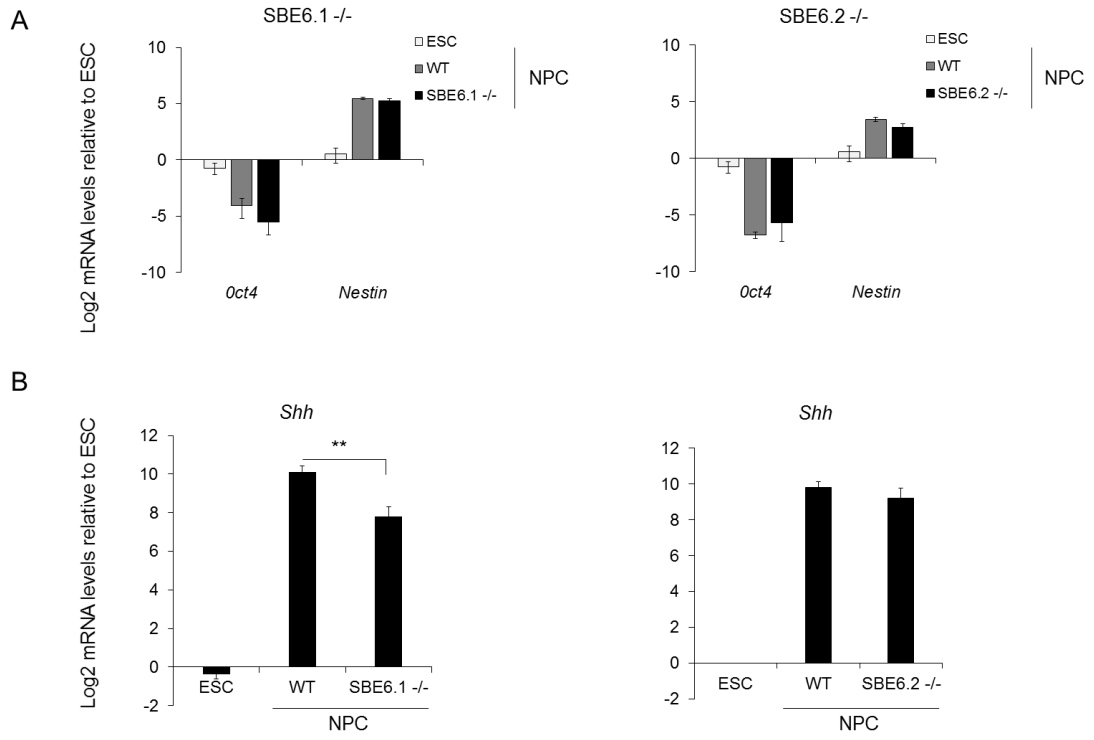

Figure Ch2- 9 Shh expression levels in neural precursor cells (NPC) derived from ESC lines with SBE6.1 or SBE6.2 deletions.

A) qRT-PCR showing mean (± s.e.m.) log2 mRNA levels of pluripotency marker *Oct4* and neural marker *Nestin* in wild-type ESC, NPC and SBE6.1$^{-/-}$ (left) or SBE6.2$^{-/-}$ (right) derived NPC, relative to *Gapdh* and normalized to ESC mRNA levels. B) qRT-PCR showing means (± s.e.m.) of relative *Shh* mRNA levels in wild type neural precursor stem cells (NPC) and in cell lines deleted for SBE6.1 (left) or SBE6.2 (right) relative to *Gapdh* and normalised to wild-type ESC. SBE6.1$^{-/-}$ data is from 5 biological replicates. SBE6.2$^{-/-}$ data are from 3 biological replicates. Statistical analysis of *Shh* expression differences was done using a one tailed student t-test. Shh are significantly reduced in NPCs derived from SBE6.1$^{-/-}$ cell lines (p-value = 0.012).

2.5 Role of SBE6.1 and SBE6.2 *in vivo*

Using the same CRISPR/Cas9 system and the same guide RNAs, with the help of Hemant Bengani a post-doc in David Fitzpatrick's team, I deleted SBE6.1 and SBE6.2 in mouse embryos by directly injecting the transcribed gRNA and Cas9 plasmid into zygotes using the same constructs as for the cell lines deletions. Hemant transcribed the gRNA plasmid into RNA and the team of the CBS-IGMM transgenic Facility co-injected the RNA and the plasmid into mouse embryos.

I recovered E10.5 and E11.5 embryos, genotyped to isolate homozygote deleted embryos and performed *in situ* hybridisation for *Shh* mRNA on those. None of the deleted embryos showed a noticeable reduction of *Shh* pattern in the brain (Figure Ch2-10). As SBE6.2 does not have a clear enhancer activity based on the zebrafish reporter assay and cell lines deletion, it is unsurprising that SBE6.2 deletion in mouse embryos does not cause disturbance on the *Shh* patterning of the brain.

Regarding SBE6.1, the analysis of embryos deletions is trickier as the zebrafish and mouse reporter assay suggested different zones of activity for the brain. Regardless of SBE6.1 forebrain or midbrain activity, other Shh enhancers such as SBE2 or SBE3 for the forebrain and SBE1 for the midbrain will be active and lead to a proper Shh pattering. Indeed, as shown in Figure Ch2-9, SBE6.1 deletion leads to a decrease of *Shh* expression in NPC but not to a total reduction, indicating that another brain enhancer is probably co-activating *Shh*. Thus it can be expected that SBE6.1 deletion in embryos would not significantly disturb *Shh* patterning. Moreover, if we consider the mouse-LacZ data to be correct, it indicates that SBE6.1 might have a very small and discontinuous domain of action, which makes it even harder to notice in an *in situ* hybridisation as the staining signal can vary from an embryo to the other and even from the left or the right side of the brain sagittal cut.

Taken together it seems that SBE6.1 enhancer activity is important in NPC differentiation in order to reach a proper *Shh* levels but other Shh brain enhancers might help to keep a proper patterning of *Shh* in the developing mouse brain.

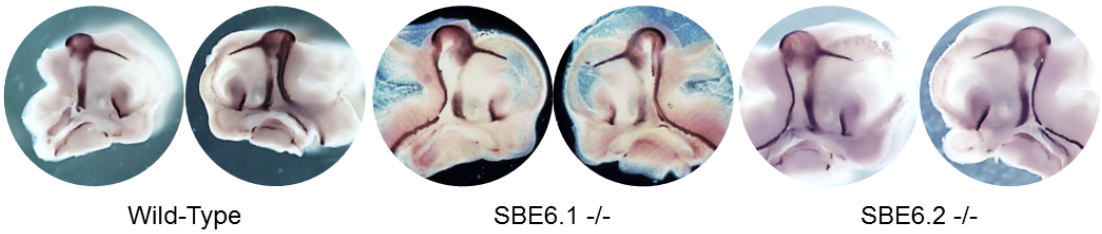

Figure Ch2- 10 *Shh in situ* hybridisation of Wild-type, SBE6.1 -/- and SBE6.2 -/- mouse embryos at stage E11.5.

Macroscope images of E11.5 mouse brain sagittal cut displaying for each case left and right side of the plane. For each condition, at least 8 embryos from 2 biological replicates (2 different placentas) were analysed.

2.6 Conclusions

By combining molecular biology techniques to identify active enhancer histone marks and *in silico* analysis to investigate conservation, accessibility and to score transcription factor binding site probability I could annotate two new putative enhancer elements for Shh that are active in the neural lineage. By then using classic developmental biology including enhancer reporter assays in zebrafish and mouse embryos, I could evaluate the function of those sequences. Interestingly only SBE6.1 was able to drive consistently expression in the forebrain of the developing zebrafish and mouse embryos. By using CRISPR/Cas9 to create with homozygote deletions I could confirm that only SBE6.1 is important for a proper *Shh* activation in NPC but does not lead to a strong *Shh* mispatterning in the developing mouse embryos indicating perhaps some level of redundancy.

It is first interesting to note that the coordinates of SBE6.1 overlap those reported for the Shh lung and gut epithelium regulatory element SLGE (Tsukiji et al., 2014), emphasizing that further analysis will be necessary to determine the critical transcription factor binding sites needed to drive enhancer function in different developmental settings. Important sequences required for enhancer function work in assemblies of transcription factor motifs (Yao et al., 2016). SBE6.1 and SLGE motifs may be intermingled but still specific to a precise tissue and stage of development, or may be overlapping to various extents. There are several examples of cis-regulatory elements with activity in multiple tissues or different cell lineages. For example, the global control region (GCR) 5' of HoxD contains regulatory information capable of driving expression in the CNS and in the limb (Spitz et al., 2003). Similarly, point mutations causing Aniridia have shown that disruption of single transcription factor binding sites can disable the ability of an enhancer to drive expression in one tissue, lens, whilst retaining enhancer activity elsewhere in development, hindbrain, diencephalon (Bhatia et al., 2013).

Regarding SBE6.2, it is curious to remark that well known SBE2, SBE3 and SBE4 enhancers are not displayed as activated using our ChIP-chip data, but this highly conserved sequence called SBE6.2 displays a tremendous gain of histone marks without driving consistently a reporter expression or being critical for *Shh* expression in NPC. This confirms the crucial need of deciphering in details the role of sequences carrying enhancer signatures.

We cannot judge an enhancer function only based on the genome-wide mapping of enhancer signatures but we need downstream developmental biology assays to decipher the function of a sequence.

The main aim of this Chapter was to inform me on which Shh brain enhancer should I focus for my thesis. Unfortunately our Chip-on-chip data did not allow me to conclude strongly on which of the known Shh-brain-enhancers are active in NPCs differentiated from ESCs. As no NPCs histone mark profiling data is publicly available, I did a broad comparison with the Shh regulatory elements in human to try to answer this aim. It is first interesting to note that our ChIP from NPC is more comparable with human neurospheres (neural stem cell culture) as it is the condition that strongly reveals enhancers mark for SBE6 position and that SBE6/SLGE appears to be similarly strongly co-localising in human (Figure Ch2-11).

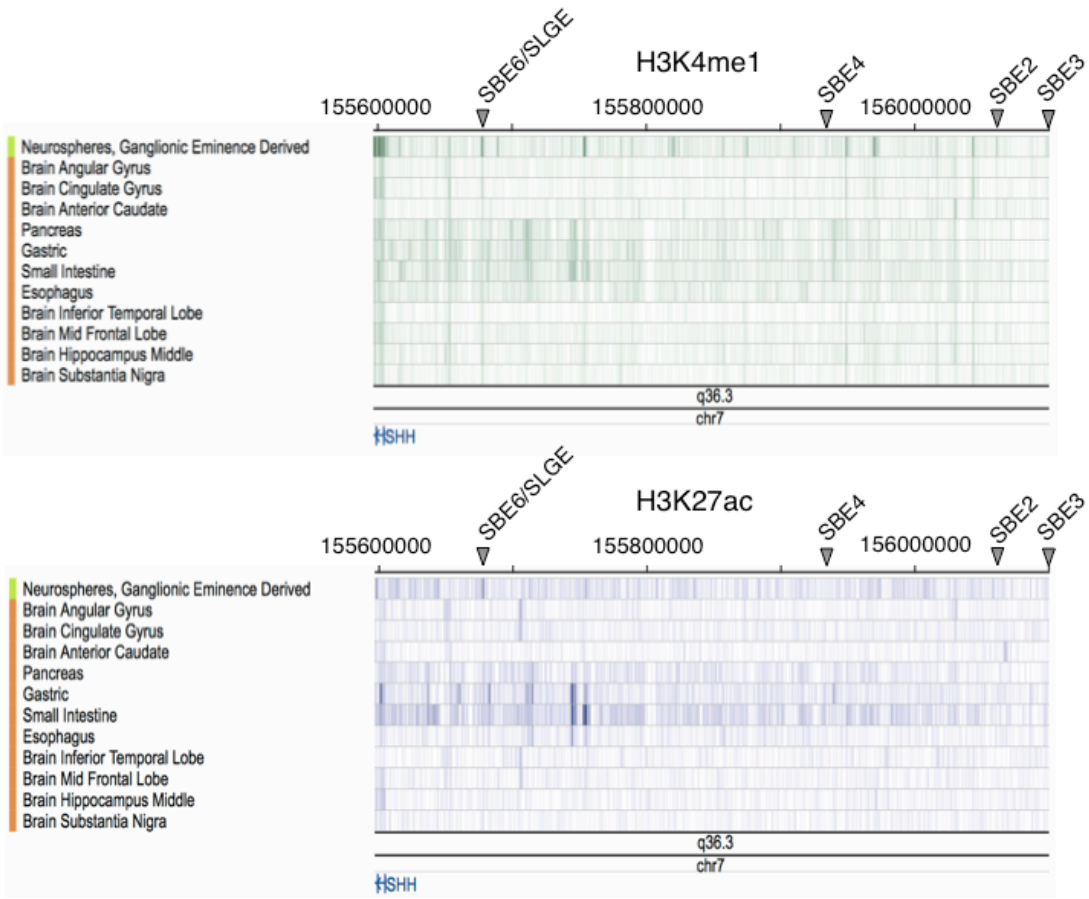

Figure Ch2- 11 Chromatin state discovery and characterization (ChromHMM) in human Shh regulatory region.

Roadmap epigenome browser view of human Shh locus on Hg19 (chr7:155595558-156100554). Arrowheads indicate corresponding enhancers in human genome.

In Chapter 2 I have revealed that SBE6.1, now named SBE6, is a cis-regulatory sequence located around 100kb upstream of *Shh* TSS that drives *Shh* expression in the mouse developing forebrain and is crucial for correct *Shh* mRNA levels in NPC.

As SBE6 seem to be the most active enhancer in our NPCs, I will follow *Shh* communication with SBE6 as well as the other known brain enhancers SBE4, SBE2 and SBE3. As SBE5 was published late in my PhD (Yao et al., 2016), I was unable to follow SBE5 precisely as our laboratory does not have a fosmid probe for SBE5, however as it is very close from the ZRS fosmid probe (≈ 30 kb), following the ZRS will give us a good indication of this SBE5 behaviour in our model.

Chapter 3: Analysing *Shh* promoter-enhancer interactions using imaging and 5C

3.1 Introduction

As discussed in the Chapter 1, despite the fact that enhancers have been studied for over thirty years, and that their central role in correct patterning of gene expression has been well established, concrete evidence for their mechanisms of action are still lacking.

Several models have been proposed to describe the interaction of enhancers with their target gene promoter. Either the enhancer is used as a secondary docking site for activators or it interacts directly with the promoter by looping of the chromatin. The first step towards understanding how an enhancer regulates a gene is to establish its mechanism of interaction, and then using this information we can subsequently question how this interaction controls gene expression.

Recent Chromosome Conformation Capture techniques have been crucial to investigate enhancer-promoter interactions. 3C and 4C have been widely used to test a precise candidate promoter-enhancer interaction or to look from a bait e.g. the promoter, which enhancer does it interact with (Ghavi-Helm et al., 2014; Montavon et al., 2011; Palstra et al., 2003; Tolhuis et al., 2002). More recently developed, Capture-C allows a focused assessment of 3D genomic interactions at high resolution (Davies et al., 2015; Hay et al., 2016; Hughes et al., 2014) and 5C and Hi-C can be used for a general mapping of a larger region or even genome-wide (Jin et al., 2013; Sanyal et al., 2012).

I decided to first use an imaging approach to investigate the mechanisms for distal enhancers and promoter interaction at the Shh locus. The main idea behind this approach is to visualize and follow enhancer-promoter communication when the gene is off and once the gene is expressed using candidate enhancers. After assessing the interaction between those enhancers and promoter by imaging I will see which of the Chromosome Conformation Capture technique would be more appropriate to further investigate the communication.

For imaging enhancer-promoter communication, fluorescence in situ hybridisation (FISH) on fixed cells has been the predominant method applied. By measuring the spatial separation between the promoter and the enhancer probes, information on the chromatin organisation can be established. Indeed, the mean-squared interprobe distances (d^2) between two FISH probes generally has a linear relationship to the genomic distance that separates them,

consistent with a random-walk polymer behaviour (van den Engh et al., 1992). Thus one can compare chromatin compaction following different conditions, between different sequences in the same cell type (Yokota et al., 1997), at different stages of differentiation or in different tissues for the same locus (Chambeyron and Bickmore, 2004; Morey et al., 2007)

Therefore FISH is an essential tool to monitor enhancer-promoter communication. By monitoring the interprobe distance between a probe at the promoter and a probe at the enhancer, I can assess how the enhancer is positioned relative to its target promoter in active cells and so gain some insight into possible models of enhancers action. If enhancer-promoter distances increase, it means that the chromatin architecture has become less tightly packaged whereas if the distance decreases it means that the enhancer comes into the vicinity of the promoter upon activation, consistent with a looping model (Chapter 1).

Previous work from Iain Williamson and Betül Hemikoglu-Balkan on the *Shh* locus in embryonic stem cells (ESC), neural progenitor cells (NPC) and limb cells revealed that *Shh* was a peculiar locus as even in non-expressing cells, the structure of the chromatin of this locus is very compact (Williamson et al., 2016). This resulted in having to deal with very small interprobe distances. Considering that the widefield microscope is limited by the diffraction of light and gives a resolution of 250nm in x-y and 600 nm in Z, I had to consider the right optical methods to image my FISH data.

3.2 Imaging of the Shh locus by light Microscopy and by Super-resolution optical microscopy

When imaged in 3D, a sub diffraction limited object – such as a FISH signal -produces a complex image that is dependent on the point spread function (PSF) of the microscope. The shape of the PSF is due to the diffraction in the microscope optical system and also depends on the fluorescent dye used for imaging, thus, it can be physically predicted and mathematically described. By setting a PSF for each dye and entering the microscope parameters into specific software, the image can be treated by deconvolution to improve the image sharpness (Figure Ch3-1A).

Mouse embryonic stem cells (ESC) that do not express *Shh* were then differentiated for seven days into neural progenitor cells (NPC) where *Shh* is expressed. Here also, I used the Sox1-GFP cell line to work on a homogeneous neural lineage. I first optimized FISH for the *Shh* locus and its brain-enhancer in 46c ESCs and derived NPCs. I compared direct and indirect labelling of the fosmid probes, using either directly fluorescently labelled d-UTP or biotin/ digoxigenin labelled d-UTP with detection using fluorescently labelled streptavidin or anti-digoxygenin antibody. I also monitored the offset created by the size of the probe and the resolution of the microscope and the most precise way of imaging data on a widefield microscope followed by image deconvolution.

By co-hybridising a Shh probe labelled separately with either FITC or Texas Red and labelled either directly or indirectly, I found that after deconvolution, direct probe labelling gave a more precise co-localisation of the two probes and less background than indirect labelling using Digoxigenin/Biotin-UTP (Figure Ch3-1A). I also established that the offset between the two Shh probe signals was smaller when *z* stacks are collected for one color – one channel – first and then switching to the other color rather than for each *z* stack (z step 120 nm) sequentially imaging the colors (Figure Ch3-1B&C). This is probably due to the very good quality of the Piezo stage used, as it memorizes its position and is able to go back

to precisely that same position with another imaging channel, whereas switching channels when at the same z plane might introduce small vibrations.

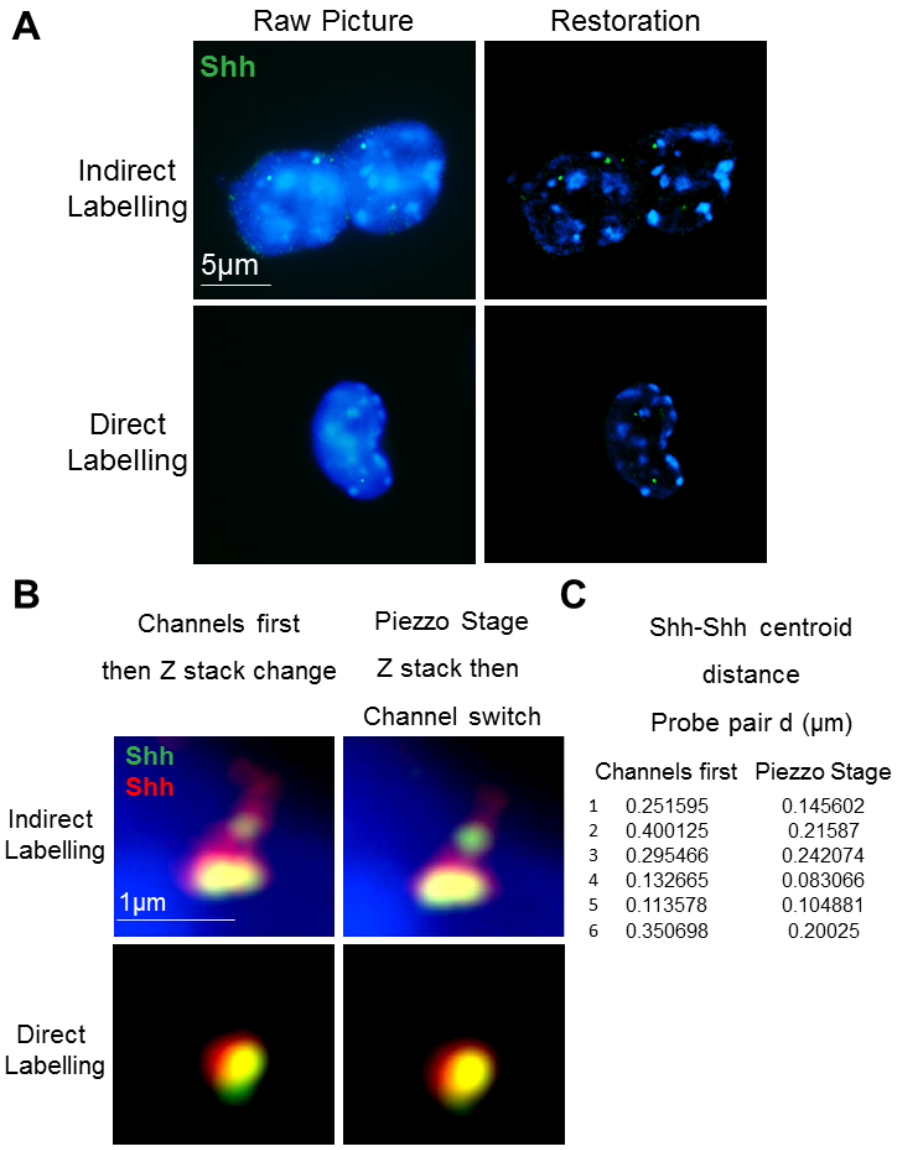

Figure Ch3- 1 Optimizing FISH image acquiring for a better resolution.

A) Raw and Deconvolved images after FISH with indirect (Digoxigenin-dUTP) and direct labelled probe (Fluorescein (5,6 Carboxyfluorescein)-dUTP). Scale bar = 5μm. B) Assessing the most accurate way of microscope acquisition by a superposition test. Imaging channel first through the depth or stopping at each z stack and switching channels of Shh labelled in two colours. Scale bar = 1μm. C) Table showing centroid distances between direct labelled Shh-red (Carboxy-X-Rhodamine (ROX)-dUTP) and Shh-green (Fluorescein (5,6 Carboxyfluorescein)-dUTP) measured from images acquired using channel first setting or Z stack first setting.

After optimizing the FISH in ESC and NPC, using direct labelling probes and sequential single channel z stack collection, I first performed FISH on ESC and differentiated NPCs to study *Shh* and its various enhancers in the region. The fosmid probes used are shown in Figure Ch3-2A.

These first results, aquired by widefield deconvolution microscopy, showed no significant changes in the distribution of inter-probe distances between Shh and control probes (Dpp6 and Control) located on the other side of Shh from its regulatory region, or between Shh and the limb-enhancer ZRS during during the differentiation of ESCs to NPCs (upper row: Figure Ch3-2B & Table 3). Surprisingly inter-probe distances between Shh and the SBE6 (SBE6 fosmid probe spans SBE6.1 and SBE6.2 enhancers) and SBE4 neural enhancer (but not SBE2/3), seem to increase in NPCs compared with ESCs (lower row: Figure Ch3-2B & Table 3). However, many of the interprobe distances imaged were below the diffraction limit of conventional widefield microscopy (usual resolution in xy ≈ 250 nm).

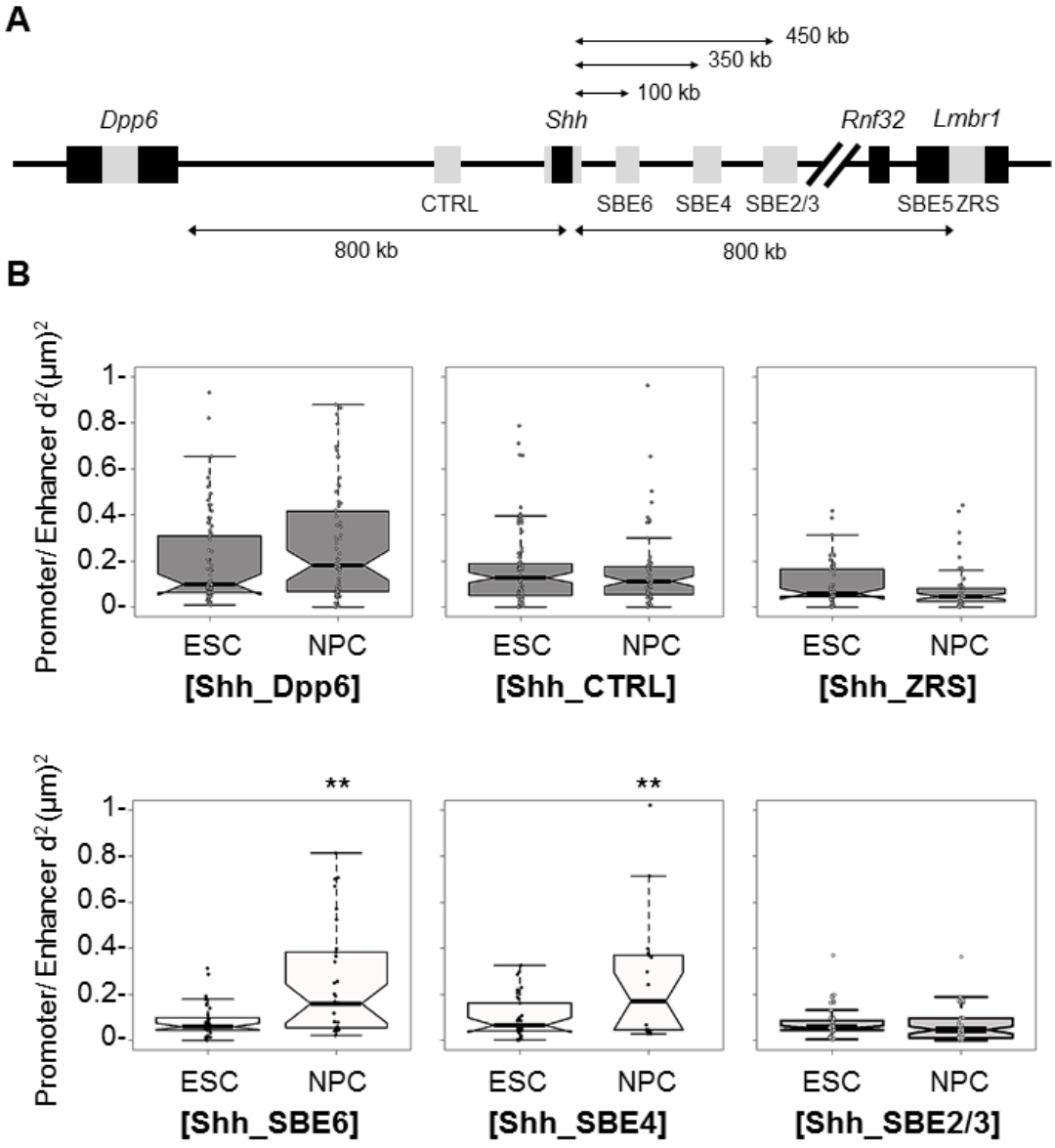

Figure Ch3- 2 3D-FISH of the Shh regulatory region by conventional wide-field microscopy.

A) Schematic Shh regulatory region map. Genes are shown as black boxes and fosmid probes as grey boxes. B) Boxplots representing squared interprobe distance (d^2,$(\mu m)^2$) around the *Shh* locus for 50-100 pairs of probes between Shh and Dpp6, ZRS, CTRL, SBE6, SBE4 and SBE2/3 fosmids. Asterisks on FISH data represent Mann-Whitney U test significance between ESC and NPC day 7 populations, ** for p-values <0.01.

ESC-NPC Day7	**Shh-Dpp6**	**Shh-CTRL**	**Shh-ZRS**
	8.40E-02	6.80E-01	5.70E-02
ESC-NPC Day7	**Shh-SBE5**	**Shh-SBE4**	**Shh-SBE2/3**
	1.32E-03	1.47E-02	1.17E-01
		1.81E-02	

Table 3 P-values from ESC and NPC Mann-Whitney U tests on FISH distances distribution acquired on a widefield microscope

Exact p-values of Mann-Whitney U tests significance between ESC and NPC day7 FISH distances distribution revealing the number of replicates.

I therefore concluded that, in order to use FISH to examine the chromosome conformation around *Shh* in relation to the newly described enhancer SBE6 (Chapter 2) and especially to ZRS or SBE2/3 that display prominent small distances, it was necessary to image the nuclei at a higher resolution than can be achieved by conventional light microscopy.

Recently, new super-resolution microscopy techniques have opened new possibilities to study nuclear ultra-structure using light down to the 100 nm resolution for xy.

Functional super resolution techniques use experimental tricks to reconstruct an image from locating precisely a signal at the time. Photo-activated localization microscopy (PALM) and stochastic optical reconstruction (STORM) use an on-off signal – a blinking due to instability in the fluorophore - to precisely locate point by point the signal and reconstruct and image. SIM is part of the "true" resolution techniques, where the physical environment is used to compute direct information from the data. By illuminating the sample with patterned light and modulating the illumination, the patterns interacts with the fluorescent probes and generate Moire fringes. From those interference patterns, extra information can be deduced and the image is computationally restored (Gustafsson, 2000). Moreover SIM has been proven to be fully compatible with FISH (Markaki et al., 2012).

Therefore with the use of 3D-SIM followed by automated image analysis software, I visualized and analysed the chromatin reorganisation upon enhancer-driven *Shh* activation using FISH

3.3 Chromatin reorganisation upon brain enhancer-driven *Shh* activation

Imaging nuclei from ESCs and differentiated neural precursor stem cells (NPC) where *Shh* is expressed by 3D-Structured Illumination Microscopy (3D-SIM) using the same set of fosmids as for the widefield imaging (Figure Ch3-3A) revealed an increased nuclear distance between *Shh* and the SBE6 (Figure Ch3-3B) and between and SBE4 upon *Shh* activation – confirming the results from wide-field deconvolution (Figure Ch3-3C & Table 4). As for conventional microscopy, nuclear distances between Shh-SBE2/3, Shh-ZRS and Shh-CTRL remain the same upon differentiation (Figure Ch3-3C & Table 4). This result is highly surprising, as distal enhancers have been extensively described as forming a loop upon activation of their cognate genes, which should be seen as decreasing interprobe distances.

My data, showing that the whole region between *Shh* and up to SBE4 undergoes a large-scale decompaction upon activation, is a new, and surprising result as it is not compatible with the looping model for enhancer-promoter communication.

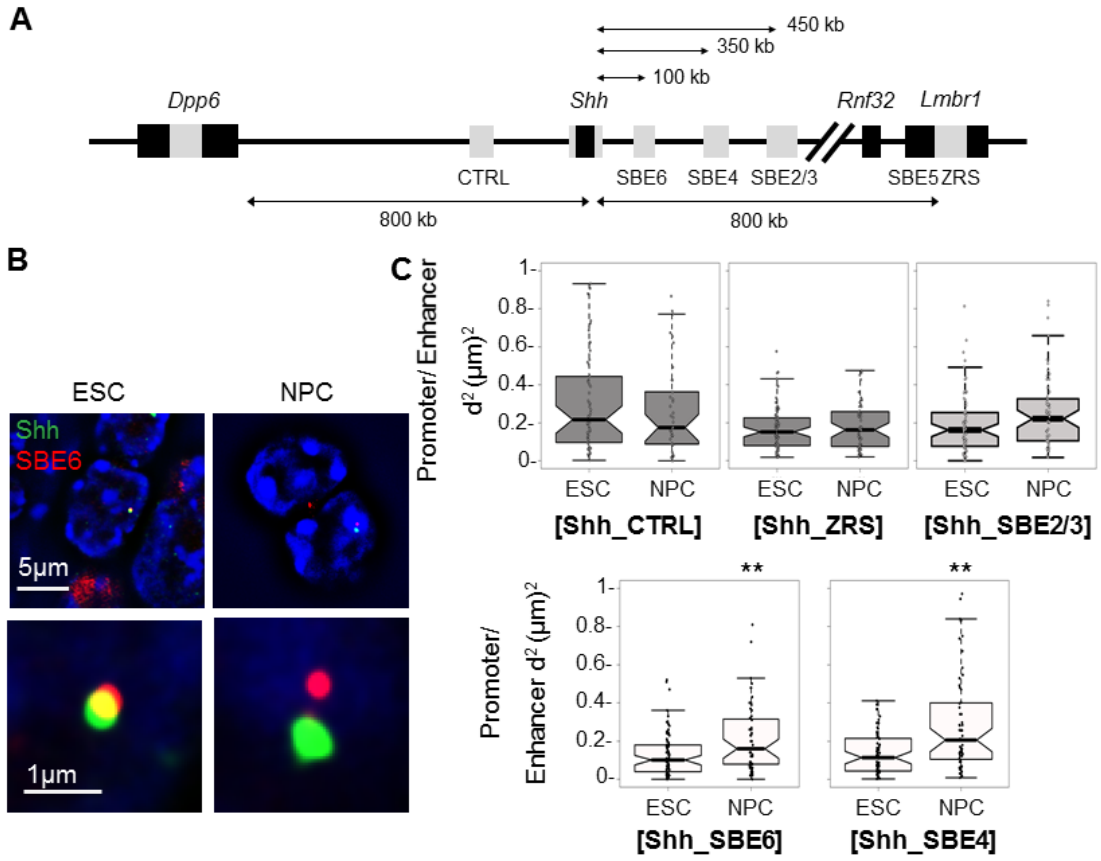

Figure Ch3- 3 3D-FISH of the Shh regulatory region using 3D-SIM.

A) Similar fosmid map as in Figure Ch3-2A. B) 3D-SIM image illustrating Shh-SBE6 distance increase in NPC after seven days of differentiation. Scales bars are 5µm top image and 1µm bottom image. C) Boxplots representing squared interprobe distances between Shh and CTRL, SBE6, SBE4, SBE2/3 and ZRS fosmids. Asterisks on FISH data represent Mann-Whitney U test significance between ESC and NPC day 7 populations, ** for p-values <0.01.

ESC-NPC Day7	Shh-CTRL	Shh-SBE2/3	Shh-ZRS	Shh-SBE6	Shh-SBE4
	4.32E-01	6.44E-01	5.94E-01	3.78E-12	3.68E-05
		5.29E-02	3.88E-01	8.21E-05	3.61E-04
		4.93E-01	6.56E-01	2.01E-02	
		9.46E-01		2.48E-03	
				8.16E-04	
				1.39E-04	
				3.78E-12	
				2.75E-03	
				1.39E-04	
				2.58E-09	
				3.33E-02	

Table 4 P-values from ESC and NPC Mann-Whitney U tests on FISH distances distribution acquired on a 3D-SIM

Exact p-values of Mann-Whitney U tests significance between ESC and NPC day7 FISH distances distribution revealing the number of replicates.

3.4 Time course of neural differentiation

It may be possible that 7 days of differentiation is too late to observe enhancer-Shh promoter looping – it may have occurred earlier in the differentiation programme. I therefore performed a differentiation time course from ESC to NPCs (day 3 to day 7). qRT-PCR shows that *Shh* mRNA levels are already increased at day 3 of differentiation and continue to increase from day 4 to 7 (Figure Ch3-4A). FISH over this same time-course showed that the Shh-SBE6 inter-probe distances increase consistently from day 4 onwards with two of three NPC day 3 replicates showing significant distance increases between (Figure Ch3-4B & Table 5).

Therefore the distance increase – chromatin unfolding - between *Shh* and the SBE6 enhancer seem to appear progressively in the population from day 4 of differentiation, but this may not be necessary for the initial stages of *Shh* expression as in day 3 cells where *Shh* is expressed at low levels, there is not always a measureable chromatin unfolding as assayed by FISH.

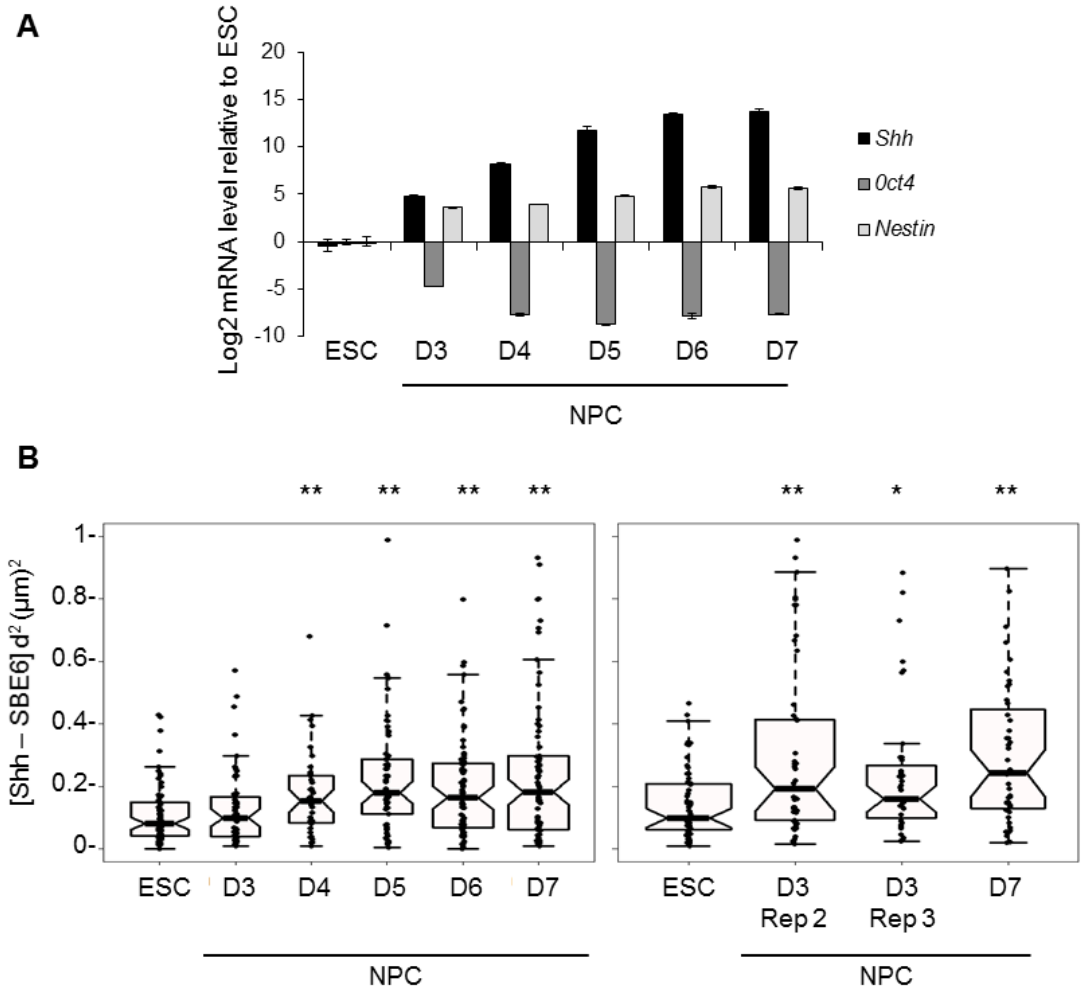

Figure Ch3- 4 Neural Differentiation characterisation.

A) qRT-PCR showing means (± s.e.m) log2 mRNA levels of *Shh*, *Oct4* and *Nestin* relative *Gapdh* and normalized to ESCs in one biological replicate. B) Left, Boxplots showing Shh-SBE6 squared interprobe distances in populations corresponding to the expression data. Right, Boxplots showing Shh-SBE6 squared interprobe distances for two other NPC day 3 biological replicates. Asterisks on FISH data represent Mann-Whitney U test significance between NPC and ESC populations, * for p-values <0.05 and ** for p-values <0.01.

	Shh-SBE6
NPC Day3	5.52E-01
	1.81E-03
	1.75E-02
NPC Day4	1.07E-02
NPC Day5	2.40E-05
NPC Day6	6.68E-04

Table 5 P-values from ESC and NPC Mann-Whitney U tests on Shh-SBE6 FISH distances distribution at various stages of neural differentiation

Exact p-values of Mann-Whitney U tests significance between ESC and NPC day 3, day 4, day 5 or day 6 Shh-SBE6 FISH distances distribution revealing the number of replicates.

However, even if the Shh-SBE6 distance increase is progressive during the differentiation time course, it might not be homogenous. Indeed, a small sub-group of cells in the population might even be looping with the promoter and trigger gene activation, when the other group is in a relaxed conformation after maybe having looped. Moreover, *Shh* expression may also not be uniform in the cell population. To test the relationship between chromatin conformation and *Shh* expression I could, in theory, perform RNA-DNA FISH where one could directly observe simultaneous transcription and correlate it or not with a Shh-SBE6 distance increase. Unfortunately *Shh* is a very small gene (≈ 8 kb) with only two introns and so detecting its nascent transcript by RNA FISH would be difficult. Indeed, a previous postdoc in the lab was unable to detect *Shh* mRNA by RNA FISH. Therefore I used another method, qRT-PCR to monitor *Shh* expression at the single cell level to assess if the $Sox1^{+ve}$ NPC population is homogenous regarding *Shh* expression or not.

3.5 Understanding Shh expression and chromatin compaction at single cell level

SHH has been shown to promote the survival of $Sox1^{+ve}$ mouse ESC derived NPCs and also to influence the determination of human ESCs toward neural lineages (Cai et al., 2008; Wu et al., 2010). Therefore I would expect that in my $Sox1^{+ve}$ NPC population, *Shh* would be expressed quite homogeneously.

In order to confirm that, I performed single cell qRT-PCR on $Sox1^{+ve}$ sorted NPCs and I could observe that the population of cells expressing *Shh* – albeit at low levels - is quite homogenous at NPCs day 3 and day 4 of differentiation with a general shift toward higher and more variable levels of mRNA production by day 7 (Figure Ch3-5A). Analysing the distribution of Shh-SBE6 FISH distances amongst the population from the Figure Ch3-4B (left) revealed that the Shh-SBE6 distance increase concerns the cell population in general with a broad shift towards higher distances with no evidence of a sub-group of cells with shorter (<200nm) distances that would be indicative of an enhancer-promoter loop. (Figure Ch3-5B). Also, without being statistically significant, at the population level in the day 3 replicate, the distribution of inter-probe distances is already slightly shifted to larger distances and becomes significant at day 4, and keeps on increasing, with in parallel decreasing p-values relative to ESCs.

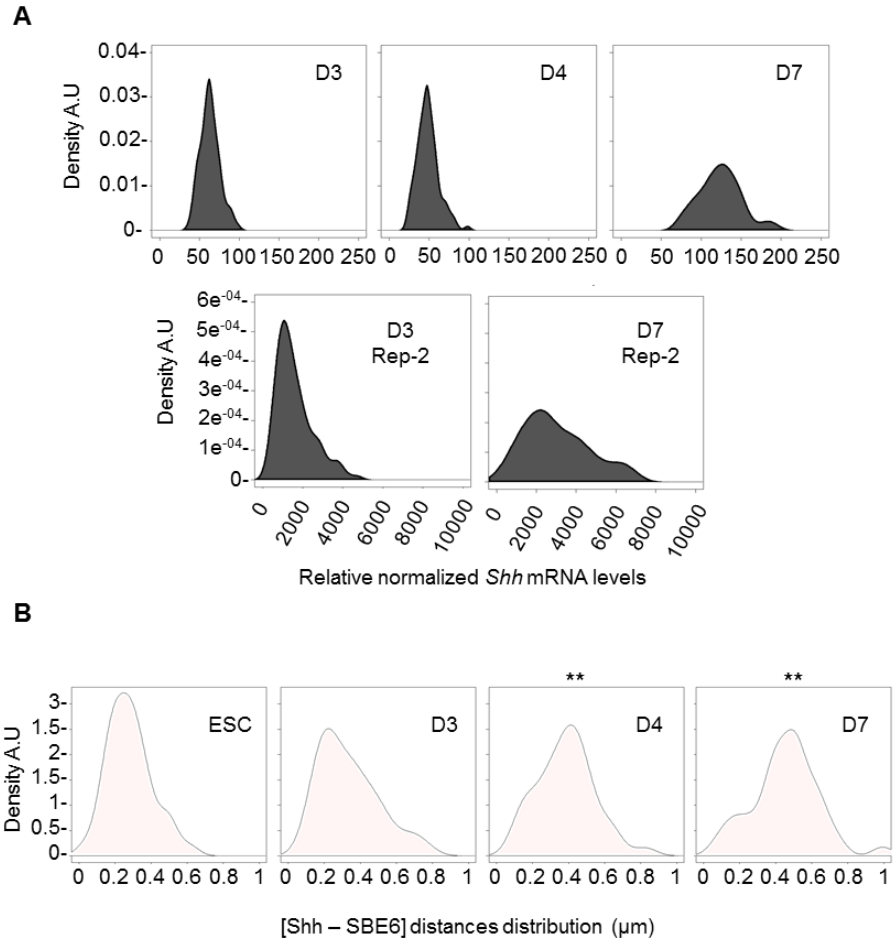

Figure Ch3- 5 NPC are a homogenous population.

A) Upper pane, Kernel density plots showing *Shh* mRNA expression of single NPC (96 cells) relative to *Gapdh* and normalised to expression in ESCs at day 3, day 4 and day 7 of differentiation. Lower panel, Kernel density plots for biological replicate showing *Shh* mRNA expression of single NPC relative to *Gapdh* and normalised to expression in ESCs at day 3 and day 7 of differentiation. B) Kernel density plots displaying Shh-SBE6 FISH distance distribution in µm in undifferentiated ESCs and at day 3, day4 and day 7 of NPC differentiation for FISH displayed in Figure Ch3-4B (left). Asterisks on FISH data represent Mann-Whitney U test significance between NPC and ESC populations, ** for p-values <0.01. Kernel density is arbitrary measured by the R software based on the size of the population and the values in order to select the appropriate binning for frequency, it is thus comparable with an arbitrary frequency unit.

Therefore the distance increase between Shh and its brain-enhancer seems to appear continuously and homogeneously in the differentiating NPC population from day 4 of differentiation. This result argues against the concept of long-range enhancer-promoter looping.

3.6 Enhancer driven chromatin unfolding

SBE6.1 -/- (SBE6-/-) derived NPCs failed to efficiently activate *Shh* expression (Figure Ch2-9). To test if the SBE6 enhancer is also required for the chromatin unfolding between Shh and SBE6 seen upon the neural differentiation of wild-type ESCs, I performed FISH on day 7 of $Sox1^{+ve}$ NPC differentiated from SBE6.1 -/- and SBE6.2 -/- cell lines described in Chapter 2 section 2.5. SBE6.1 -/- derived NPCs failed in contrast to recapitulate the Shh-SBE6 nuclear distance increases (Figure Ch3-6A & Table 6), correlating with the loss of their *Shh* enhancer activity described in Chapter 2. Whereas SBE6.2 -/- NPCs, which still activate *Shh* (Figure Ch3-6), do recapitulate the Shh-SBE6 distance increase seen in wild-type NPCs (Figure Ch3-6B & Table 6), meaning that the loss of chromatin unfolding is not due to a deletion of any genomic region.

Taken together these results suggest that the chromatin unfolding between Shh and SBE6 observed by FISH is linked with Shh-Brain-Enhancer SBE6.1 activity. As also highlighted by the qRT-PCR and FISH data, *Shh* seems to be expressed at basal levels in day 3 NPC and in NPC lacking SBE6.1, and chromatin unfolding of this locus seems to be linked not with this basal level of *Shh* expression, but with the enhanced expression of *Shh* mRNA.

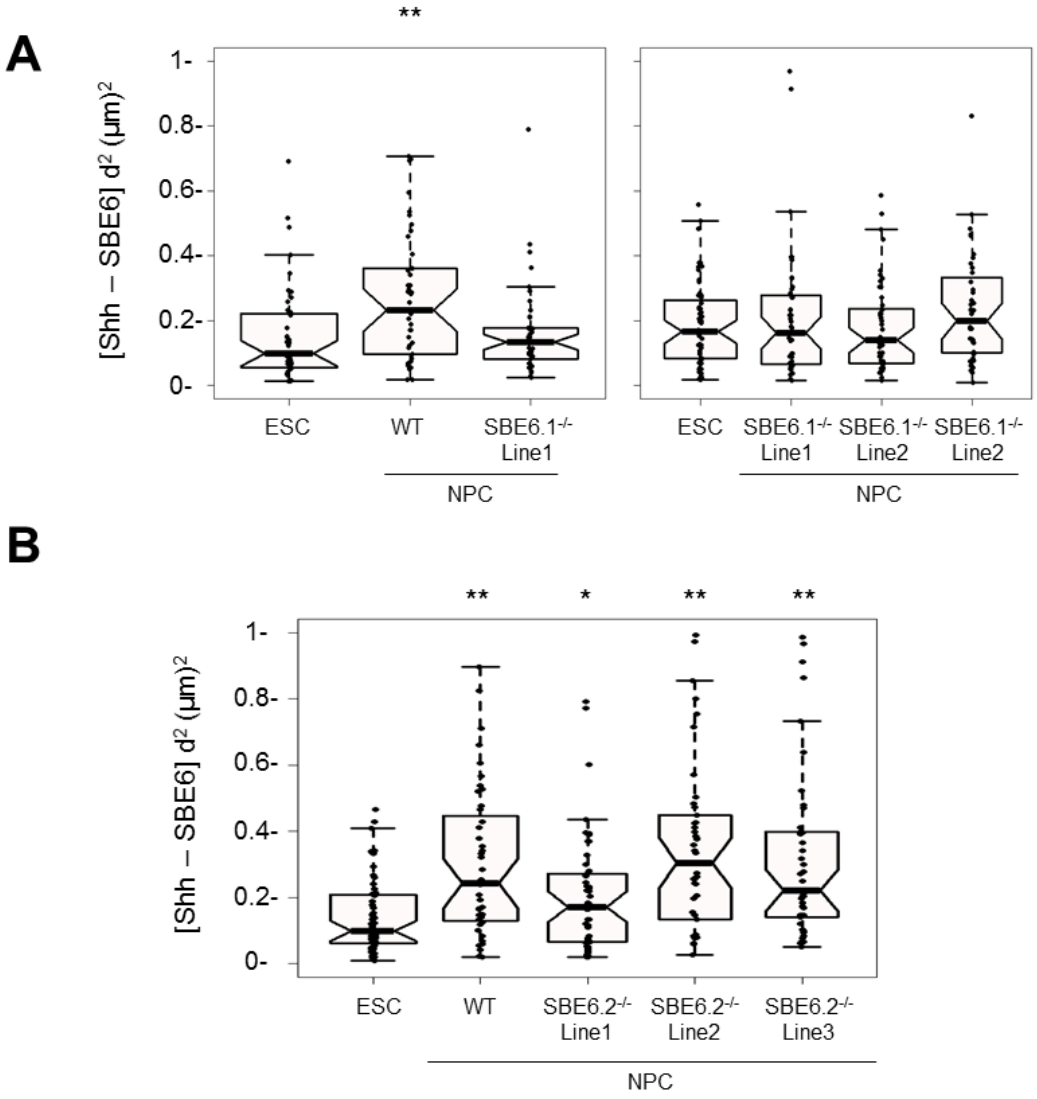

Figure Ch3- 6 Enhancer driven chromatin unfolding.

A) Boxplots representing Shh-SBE6 distances FISH data on NPC derived from three independently generated SBE6.1-/- cells (SBE6-/-) and one second biological replicate for SBE6.1-/- Line 1. B) Boxplots representing Shh-SBE6 distances FISH data on NPC derived from three independently generated SBE6.2-/- cells. Asterisks on FISH data represent Mann-Whitney U test significance between NPC and ESC populations, * for p-values <0.05 and ** for p-values <0.01.

	Shh-SBE6		Shh-SBE6
NPC SBE6.1 -/-	3.06E-01	**NPC SBE6.2 -/-**	1.01E-02
	9.70E-01		3.47E-06
	6.62E-01		9.50E-06
	1.97E-01		

Table 6 P-values from ESC and SBE6 KO NPC Mann-Whitney U tests on Shh-SBE6 FISH distances distribution

Exact p-values of Mann-Whitney U tests significance between ESC and SBE6.1-/- or SBE6.2-/- NPCs Shh-SBE6 FISH distances distribution revealing the number of replicates.

3.7 Chromatin conformation capture carbon-copy (5C)

As a complement to FISH, chromatin organization studies have recently been undertaken using the Chromosome Conformation Capture (3C) techniques and its derivatives (4C, 5C, 3C-Seq, capture-C, HiC) (Dekker et al., 2002; Dostie et al., 2006; Hughes et al., 2014; Lieberman-Aiden et al., 2009; Simonis et al., 2006; de Wit and de Laat, 2012). Ligation frequencies are measured between two distal sequences that have been captured together in the same formaldehyde cross-linked complex, and are used to infer the spatial proximity of the two sequences in the fixed nuclei. These methods have been used to identify topologically associated domains (TADs) that are separated by boundaries often enriched for CTCF, cohesin and other architectural proteins (Dixon et al., 2012; Jin et al., 2013; Nora et al., 2012; Rao et al., 2014; Seitan et al., 2013; Sofueva et al., 2013; Zuin et al., 2014). However, FISH and 3C methods do not always produce concordant results and so both methods should be validated against each other in order to draw firm conclusions (Williamson et al., 2014, 2016).

I therefore performed Chromatin Conformation Capture Carbon Copy (5C) in ESC and NPC on the *Shh* region to see if imaging data could be transposed in terms of the interaction frequencies of the chromatin and also if *Shh* is strongly interacting i.e. looping, with an element that I have not considered in my imaging studies as Chapter 2 revealed many more putative Shh-brain-enhancer not yet annotated.

The appearance of a TAD boundary that starts before *Shh* and ends before *Nom1* seems to be consistent between ESC and NPC, and consistent with previously published HiC data from mouse ESC (Dixon et al., 2012; Williamson et al., 2016) (Figure Ch3-7A). Moreover FISH and 5C observation indicate that Shh TAD seems to be consistently closed by a "loop" formed with sequence spanning the ZRS and *Shh* both in cells expressing and non-expressing *Shh* (Shh and ZRS are in a constrained nuclear space with 3D distances much smaller than what is expected from their genomic position, however specific co-localisation occur only in expressing cells) (Williamson et al., 2016) (Figure Ch3-2B, Ch3-3C, Ch3-7A & Ch3-8A).

Analysing the 5C heatmaps with a first degree of zooming in the Shh regulatory region revealed that the sequence spanning *Shh* seems indeed to interact with ZRS but do not seem to strongly interact more with SBE5 or ZRS in NPC, which is more precisely seen by FISH (Figure Ch3-7B & Ch3-8B, Ch3-3C).

Zooming in and comparison between ESC and NPC heatmaps also reveal the loss of a strong interaction between a region upstream *Shh* and another sequence near Rnf32 on the far right (Figure Ch3-7C). This strong interaction corresponds to two CTCF sites that are interacting in ESC. By CTCF Chip performed by Betül Hemikoglu-Balkan, we could see that the CTCF peak upstream *Shh*, indeed disappeared upon differentiation into NPC (Figure Ch3-9) therefore explaining the loss of interaction between this sequence and the other CTCF peak on the far right.

This loss of CTCF interaction could be linked to the chromatin opening up upon NPC differentiation that I have seen by FISH. Further comparison of the profiles could highlight also a possible broad decompaction surrounding SBE6 and SBE4 sequences pointed by arrows and arrowhead (Figure Ch3-7C & Ch3-8C). Most importantly, this first set of 5C data shows that *Shh* is not interacting with a missed enhancer or any other sequence upon activation in NPC (Figure Ch3-7 & Ch3-8).

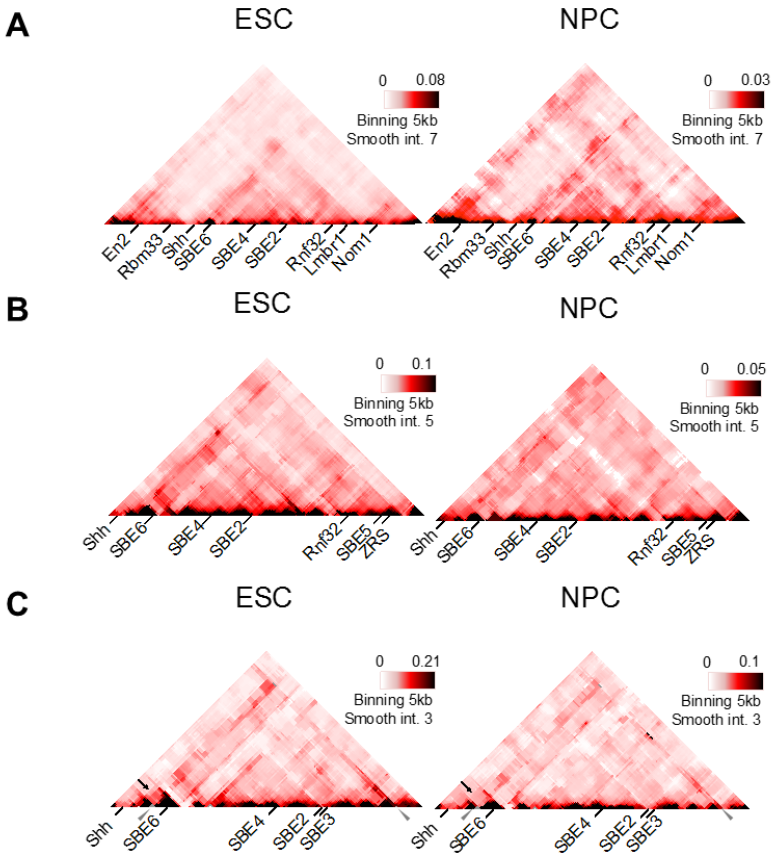

Figure Ch3- 7 Chromatin conformation capture carbon-copy (5C) heat-maps of the Shh regulatory region in ESC and NPC.

A) 5C heatmaps of ESC and two biological replicate of NPCs beyond Shh regulatory region with 5kb binning and smoothing with the median of 7 surrounding interactions frequencies revealing a TAD of the Shh regulatory region. 5C heatmap of each biological replicates has been created from the average reads from sequencing technical replicates. B) 5C heatmaps of the Shh regulatory region (chr5:28750000-29750000) with 5kb binning and smoothing with the median of 5 surrounding interactions frequencies. C) Zoomed 5C heatmaps on the Shh regulatory region (chr5:28750000-29450000 in ucsc mm9) with 5kb binning and smoothing with the median of 3 surrounding interactions frequencies. Black arrows highlights the differences with ESCs and grey arrowhead show CTCF sites.

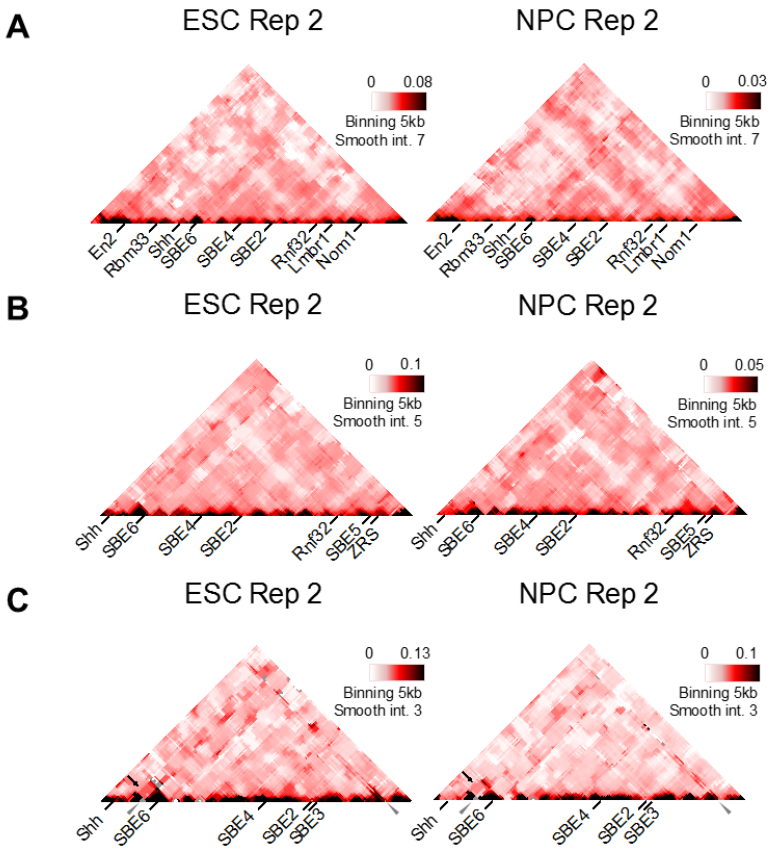

Figure Ch3- 8 Chromatin conformation capture carbon-copy (5C) heatmaps of the Shh regulatory region in ESC and NPC for biological replicates.

A) 5C heatmaps of ESC and NPC biological replicates beyond Shh regulatory region with 5kb binning and smoothing with the median of 7 surrounding interactions frequencies revealing a TAD of the Shh regulatory region. 5C heatmap of each biological replicates has been created from the average reads from sequencing technical replicates. B) 5C heatmaps of the Shh regulatory region (chr5:28750000-29750000) with 5kb binning and smoothing with the median of 5 surrounding interactions frequencies. C) Zoomed 5C heatmaps on the Shh regulatory region (chr5:28750000-29450000 in ucsc mm9) with 5kb binning and smoothing with the median of 3 surrounding interactions frequencies. Black arrows highlights consistent differences and grey arrowheads points at CTCF sites.

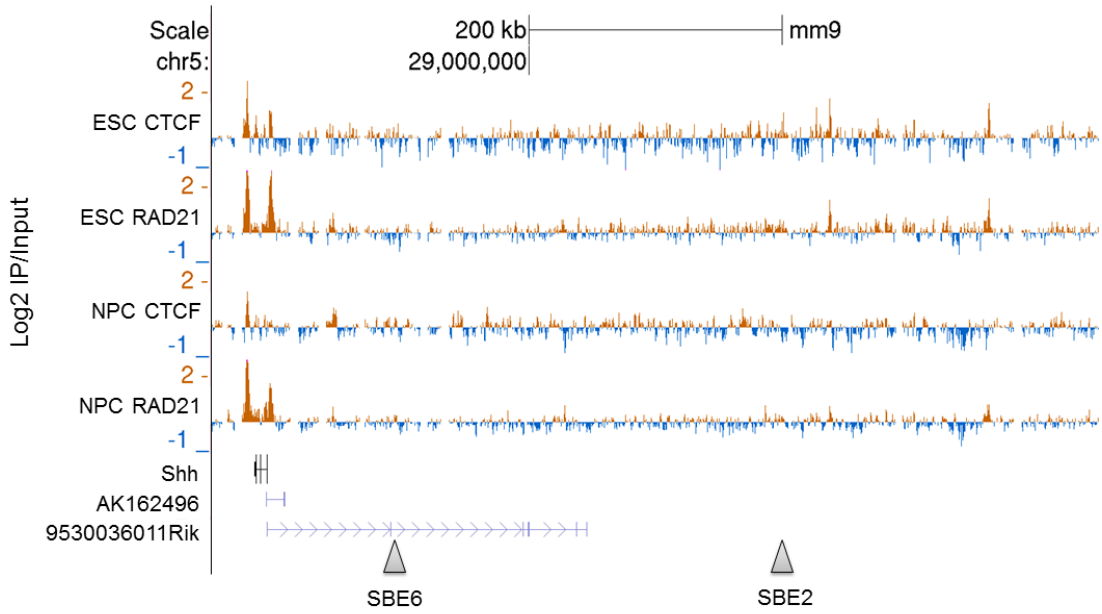

Figure Ch3- 9 CTCF and Rad21 ChIP-chip.

UCSC Genome browser view of chr5:28,750,000-29,400,000 mm9 displaying log2 CTCF and Rad21 ChIP relative to input on Shh regulatory region on ESCs and NPCs and after Tales-Vp64 transfection in ESC, three biological replicates combined. Arrowheads represent enhancer location SBE6, SBE2.

Appearance or disappearance of strong direct enhancer-promoter interaction can usually be detected by comparing enhancer-promoter ligation products in different conditions using the Chromosome Conformation Capture (3C) technique and its derivatives (4C, 5C, Hi-C etc.) (Jin et al., 2013; Montavon et al., 2011; Palstra et al., 2003; Tolhuis et al., 2002). However establishing any other type of further analysis such as general chromatin conformation changes are much harder to base on these assays. First, as these assays only give an average interaction frequency in the population, if the interaction is not constrained (precise interaction between sequences versus a broad sequence with a broad sequence), such as the distance increase observed, the distances are distributed in a wide range of possibilities which will not be represented by chromosome conformation capture techniques. Overall it appears that "C" techniques are best able to detect strong and stable interactions that will not be diluted amongst the population. As the Shh-SBE6 distances distribution spans a wide range of states (Figure Ch3-5B), we could presume that the interaction between *Shh* and its brain-enhancer upon activation will hardly been seen on 5C heat-maps as the distance is not stable enough to lead to a consistent ligation product.

3.8 Conclusions

This Chapter showed that using FISH and super-resolution microscopy, upon neural activation, *Shh* and its brain-enhancer seem not to follow the popular enhancer-promoter looping model. On the contrary, the locus shows a large enhancer-dependent chromatin unfolding that is continuous and homogenous amongst the population in differentiating NPCs.

It remains possible that I do not observe a loop due to limitations in the 3D SIM technique at this highly compact locus. In the following chapter I go on to explore this possibility to confirm the unexpected observation of enhancer mediated chromatin unfolding.

Chapter 4: Using a synthetic biology approach to study enhancer-promoter interaction

4.1 How to measure a chromatin loop?

As described in the previous chapters, direct enhancer-promoter interaction has been traditionally established by imaging (Amano et al., 2009; Ronshaugen and Levine, 2004; Schoenfelder et al.) or "C" techniques (Deng et al., 2012; Ghavi-Helm et al., 2014; Vernimmen et al., 2007). However, in our highly compact locus, both techniques did not reveal a looping of the chromatin that would indicate a direct interaction between *Shh* and a brain-enhancer.

In order to confidently refute the looping model for this locus in the neural context, I would need to be able to visualise a chromatin-looping event as a positive control.

Using a synthetic biology approach, recent chromosome engineering experiments at the *β-globin* locus have established that imposing a direct interaction between an enhancer and its cognate promoter by chromatin looping is sufficient to promote gene expression in erythroid cells provided that the correct transcription factors are available (Deng et al., 2012, 2014). By forcing the targeting of the self-association domain (SA) of LIM domain-binding protein 1 (LDB1) – a key transcription factor- via an engineered zinc finger protein to the LCR, *β-globin* expression was upregulated and a chromatin loop viewable by 3C (Figure Ch4-1). Recent follow-up experiment suggests that the enhanced globin expression is due to an increased frequency of transcriptional bursts and not to an increase in burst size, revealing a highly dynamic interaction (Bartman et al., 2016).

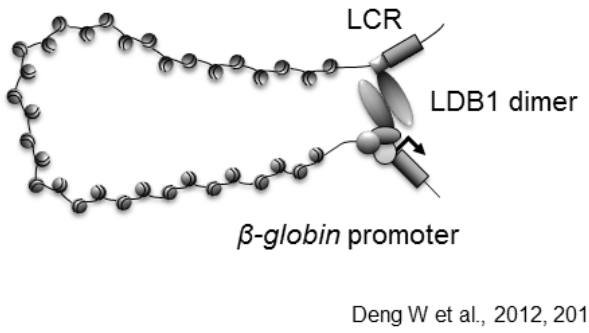

Figure Ch4- 1 Schematic representation of the LDB1 dimerization.

LDB1 dimerization induces a loop via a direct interaction of the LCR and the *β-globin* promoter and that is enough to induce gene activation (Deng et al., 2012, 2014).

4.2 Using a synthetic approach based on TAL effectors

Transcription activator-like effector (TALE) proteins are DNA-binding proteins (from a plant pathogen) that contain a tandem repeat sequences called repeat-variable di-residues (RVDs) where a simple cipher governs DNA binding of the repeat residues (Moscou and Bogdanove, 2009).

Thus TALEs can be engineered to bind any DNA sequence required with high levels of specificity. These DNA-binding proteins can then be fused to any factor to force its recruitment to a sequence of interest. This approach has been used to activate transcription or to unfold chromatin structure at specific gene promoters in mouse ES cells (Therizols et al., 2014)

I designed a TALE to bind the *Shh* promoter (tShh) and two of the Shh-Brain-Enhancer (tSBE6 for SBE6.1 and tSBE2). These were then fused to the self-association domain (SA) of LIM domain-binding protein 1 (LDB1) (Figure Ch4-2A).

By introducing two TALEs-LDB1 into ESCs, Shh and the SBE6/SBE2 enhancer could interact directly via dimerization of the LDB domain. FISH observation showed that the looping did indeed occur upon tShh-LDB1 and either tSBE6-LDB1 or tSBE2-LDB1 co-transfection since the inter-probe distance decreased significantly compared to the eGFP transfection controls (Figure Ch4-2B & Table 7).

However, simply inducing looping did not lead to *Shh* expression compared to the negative control, eGFP transfected cells (Figure Ch4-2C). This is not surprising since the action of LDB1 in activating *β-globin* still requires the presence of other erythroid transcription factors – i.e. GATA1.

Proceeding to a four-color 3D-FISH, to probe the sequence in between the direct-interaction, I could confirm TALE-LDB1 co-transfection do not trigger a general aggregate of the chromatin (Figure Ch4-3A, B & Table 7).

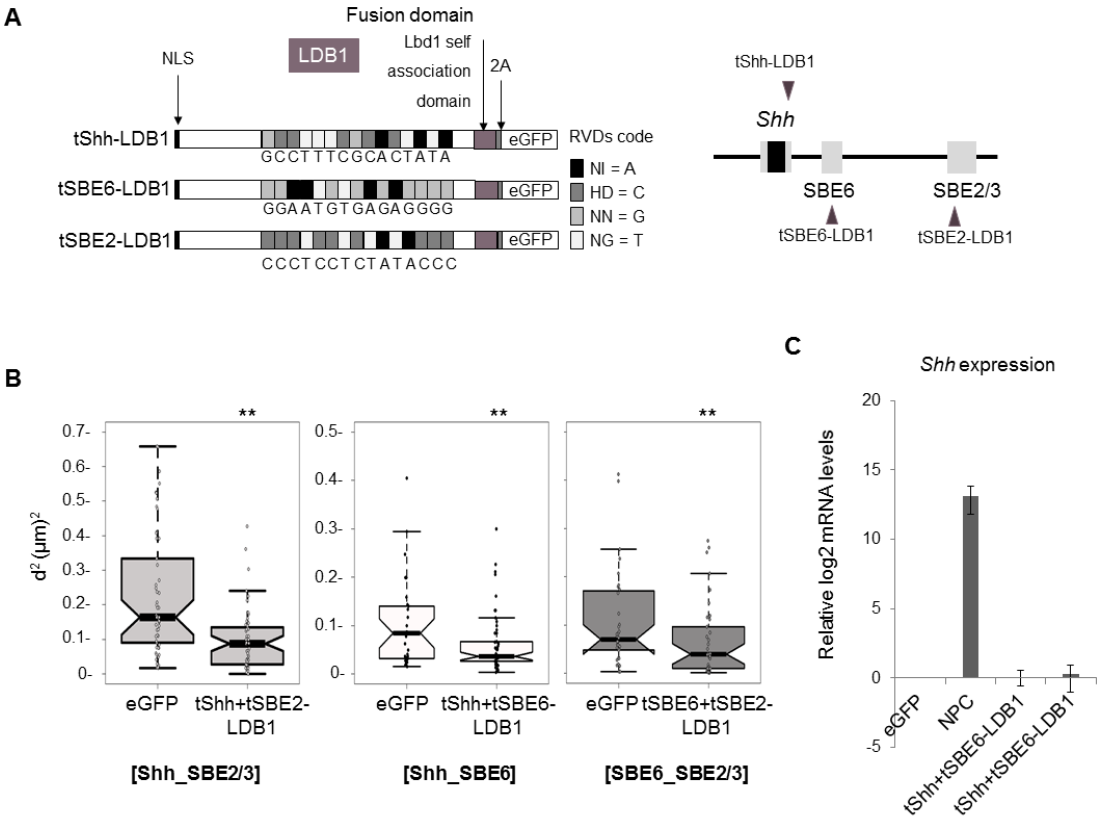

Figure Ch4- 2 Tale-LDB1 induces direct interaction between Shh and its Brain-enhancers.

A) Schema representing Tale-LDB1 targeting sequence and constructs with map of targeting sites. Self-cleaving (2A) peptide allows parallel expression of eGFP and isolation of TAL-expressing cells by fluorescence-activated cell sorting (FACS). RVD, repeat variable diresidue is displayed with one-letter abbreviations for amino acids. B) Tale-LDB1 forcing the looping between Shh and SBE6 or SBE2. Boxplots representing Shh-SBE2/3 FISH squared interprobe distances after tShh+tSBE2-LDB1 transfection. Right, Boxplots representing squared interprobe distances for Shh-SBE6 or SBE6-SBE2/3 after tShh+tSBE6-LDB1 or tSBE6+tSBE2-LDB1. C) qRT-PCR displaying means (± s.e.m) log2 mRNA levels of *Shh* relative to *Gapdh* after direct interaction in ESC between Shh and its Brain-enhancer SBE6 or SBE2 for three biological replicates normalized to eGFP plasmid transfection in ESCs as negative control.

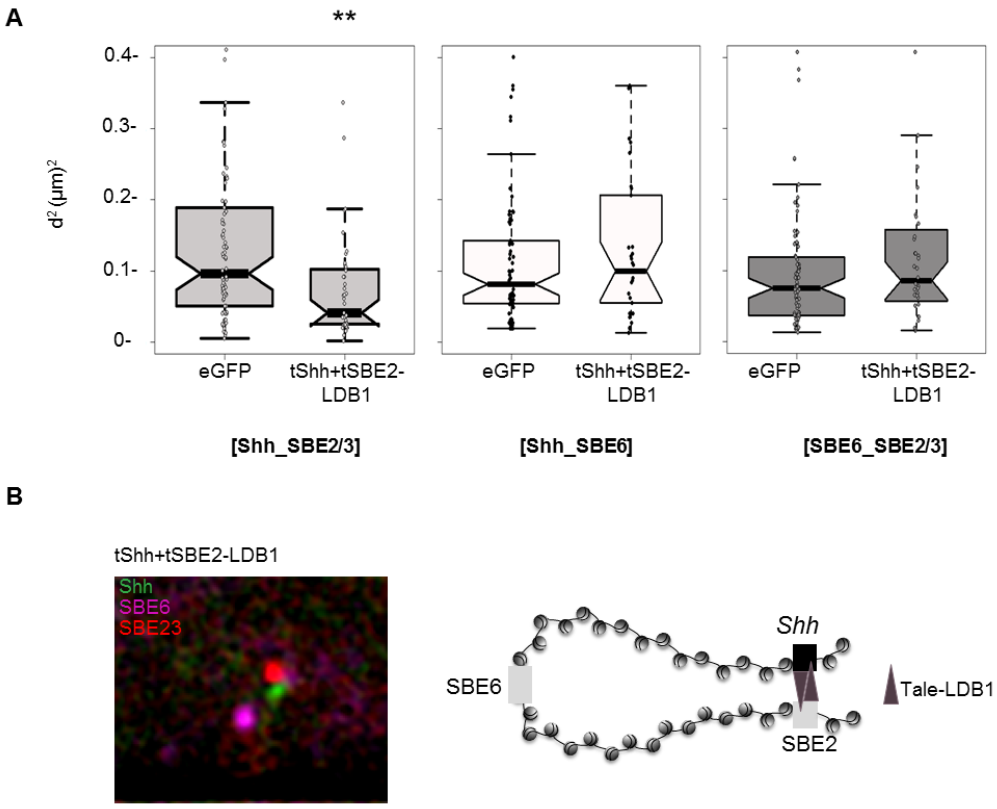

Figure Ch4- 3 Tale-LDB1 triggers a direct interaction between Shh and SBE2 and a looping out of the intervening chromatin.

A) Boxplots representing squared interprobe distances of 4 colour FISH for Shh-SBE2/3, Shh-SBE6, SBE6-SBE2/3 after tShh+tSBE2-LDB1. Asterisks on FISH data represent Mann-Whitney U test significance between Tale-LDB1 and eGFP transfected populations, ** for p-values <0.01. B) 4-Color FISH representing that Tale-LDB1 co-transfection leads to a precise dimerization and Illustration of loop formation with Tale-LDB1.

	Shh-SBE6	Shh-SBE2/3	SBE6-SBE2/3
tShh+tSBE2- LDB1	0.1666	4.3E-06	0.8953
	0.4571	0.01661	0.1976
		0.0003	
tShh+tSBE6- LDB1	0.0217		
tSBE6+tSBE2- LDB1			0.01754

Table 7 P-values from eGFP and Tale-LDB1 transfected ESC Mann-Whitney U tests on FISH distances distribution

Exact p-values of Mann-Whitney U tests significance between eGFP and Tale-LDB1 transfected cells revealing the number of replicates.

Therefore FISH is able to detect chromatin loops if they are deliberately generated. This supports my conclusion that during normal *Shh* activation in neural differentiation the simple enhancer-promoter looping model does not apply.

These data highlight that in order to understand enhancer-promoter interactions individual loci need to be specifically investigated, and that there is no universal mechanism of enhancer action at a distance.

4.3 Creating an enhancer bypass in ESC

TALE fused to transcription factors or repressors have been extensively used in recent years to modulate artificially gene expression. TALE-activators and TALE-repressors have been successfully used to activate or repress gene expression when targeted to their respective enhancer elements by recruiting Vp64 – Four repeats of the small viral acidic protein Vp16 that can strongly activate genes when bound to them (Therizols et al., 2014; Zhang et al., 2011) - or recruiting KRAB protein to a specific promoter, the targeted gene could be artificially repressed (Cong et al., 2012).

As transcription and the associated supercoiling is known to remodel large chromatin domains and to change chromatin compaction (Matsumoto and Hirose, 2004; Naughton et al., 2013), it is possible that the chromatin reorganisation I had seen using FISH during NPC differentiation just occurs passively as the consequence of *Shh* activation. Targeting Vp64 directly to the Shh promoter, I should be able to activate *Shh* expression without the involvement of any Shh enhancers. By fusing tShh to Vp64 (tShh-Vp64) (Figure Ch4-4A) it is possible to artificially activate *Shh* in ESC at similar levels as found in NPC day seven without triggering differentiation (Figure Ch4-4B&C). However, expression of tShh-Vp64 did not lead to the same nuclear distance increases as observed in NPC during endogenous *Shh* activation (Figure Ch4-5A Left & Table 8).

I then activated the Shh-brain-enhancer SBE6 and SBE2 using tSBE6-Vp64 and tSBE2-Vp64 alongside a co-transfection of both tSBE6-Vp64+tSBE2-Vp64. All succeeded to activate *Shh* in ESC, though not to the same extent as with tShh-Vp64 (Figure Ch4-4B).

Shh expression levels show that combinations of Tale-Vp64 targeted to *Shh* regulatory region can activate *Shh* without inducing neural differentiation as *Oct4* and *Nestin* remains at similar levels as in ESC (Figure Ch4-4C).

However, none of the single Tale-transfections led to the same nuclear distance increase as seen during the endogenous activation of *Shh* during differentiation.

Surprisingly, only the co-activation of SBE6 and SBE2 via the co-transfection of tSBE6-Vp64+tSBE2-Vp64 led to a significant distance increases between Shh and SBE6 (Figure Ch4-5A & Table 8).

The FISH interprobe distance between Shh and SBE2/3 after Tale-Vp64 transfections did not show any nuclear distance change which shows that this chromatin unfolding is similar to that observed during endogenous neural differentiation (Figure Ch4-5B). I could also control that this effect is specific to Vp64 recruitment and not just due to Tale protein targeting. Recruiting two Tales-Deltas, i.e. Tales without any fusion domain, does not recapitulate the distance increase (Figure Ch4-5C).

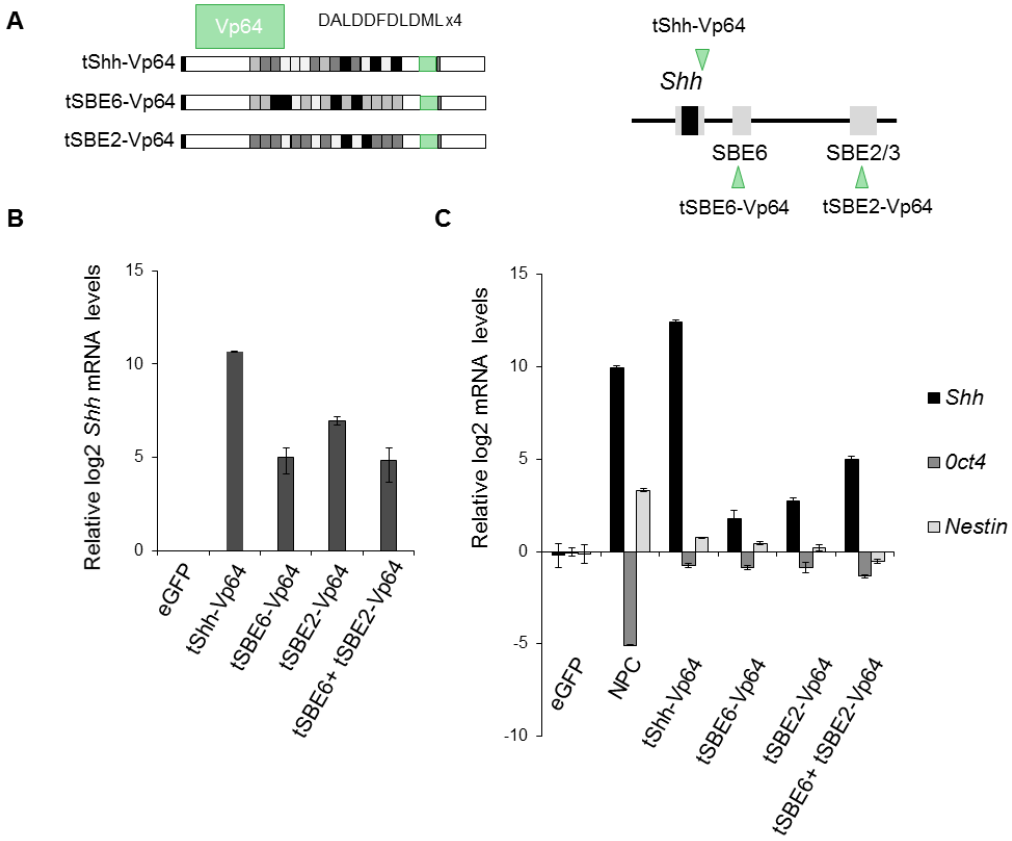

Figure Ch4- 4 Shh Enhancer Bypass in ESC using Tale-Vp64 1.

A) Schema representing Tale-Vp64 targeting sequence and constructs with map of targeting sites. B) qRT-PCR displaying means (± s.e.m) log2 mRNA levels of *Shh* relative to *Gapdh* after activation in ESC with Tale-Vp64 for three biological replicates normalized to eGFP transfection. C) qRT-PCR displaying means (± s.e.m) log2 mRNA levels of *Shh*, *Oct4* and *Nestin* relative *Gapdh* and normalized to ESCs in one biological replicate.

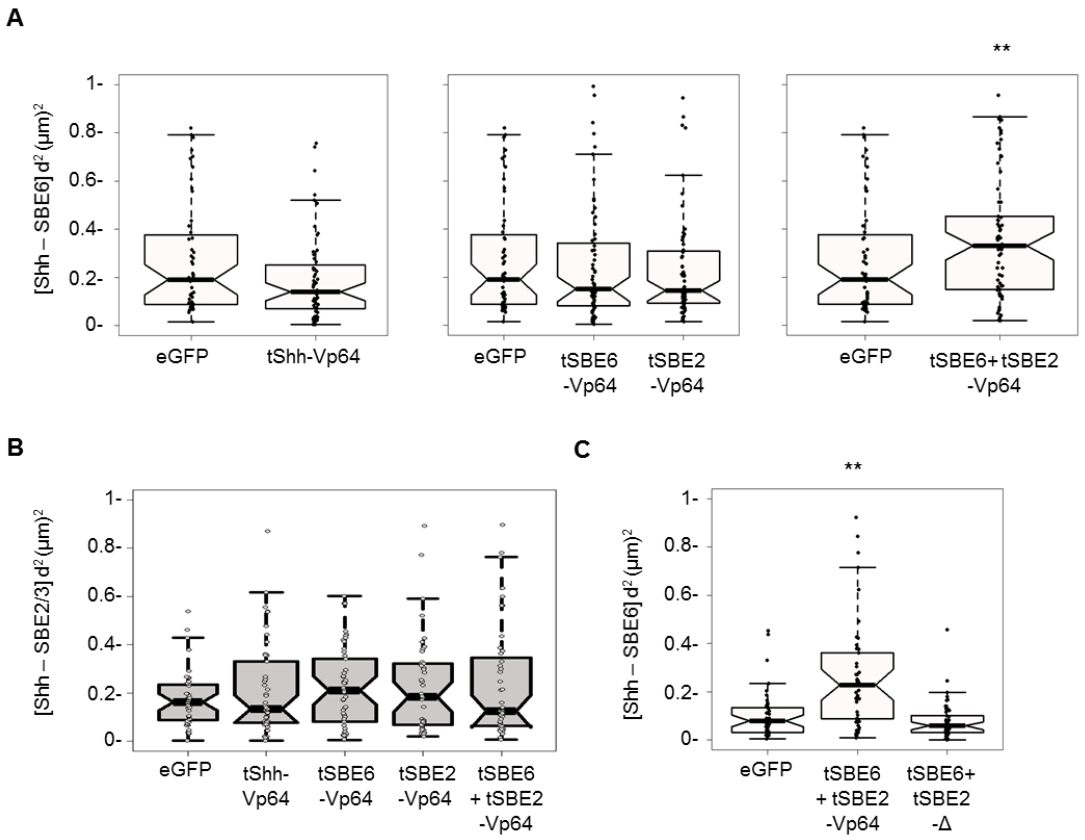

Figure Ch4- 5 Shh Enhancer Bypass in ESC using Tale-Vp64 2.

A) Boxplots representing Shh-SBE6 squared interprobe distances after Tale-Vp64 transfections. B) Boxplots representing Shh-SBE2/3 squared interprobe distances after Tale-Vp64 transfections. C) Boxplots representing Shh-SBE6 squared interprobe distances after Tale-Vp64 and Tale-Delta (no protein fused) transfections.

	Shh-SBE6	Shh-SBE2/3
tShh-Vp64	5.10E-02	3.11E-01
	3.27E-01	4.25E-01
tSBE6-Vp64	3.64E-01	9.75E-01
	2.91E-01	
tSBE2-Vp64	4.47E-01	7.68E-01
tSBE6+tSBE2-Vp64	5.45E-12	4.10E-01
	3.85E-05	
	1.33E-02	
	6.50E-08	
	2.73E-06	
	1.93E-11	
	1.95E-06	
	4.52E-02	
tSBE6+tSBE2-Δ	2.28E-01	

Table 8 P-values from eGFP and Tale-Vp64 transfected ESC Mann-Whitney U tests on FISH distances distribution

Exact p-values of Mann-Whitney U tests significance between ESC and Tale-Vp64 transfected ESCs revealing the number of replicates.

To determine whether the activators position matters, that is to say if only activating a known enhancer element was required, I also designed a new Tale for a site located between Shh and SBE6 (50kb from Shh) that does show any evidence of enhancer function (H3K4me1/H3K27ac) during ES-NPC differentiation (Figure Ch4-6A). This Tale-Vp64 (called NE – for Non-Enhancer) was able to activate *Shh* expression alone or together with tSBE6-Vp64 or tSBE2-Vp64 (Figure Ch4-6B). However, once again none of the single Tale-transfections led to the same nuclear distance increase as seen during the endogenous activation of *Shh*. Only the co-targeting of multiple Tale-Vp64s to the Shh regulatory region (tNE-Vp64+tSBE6-Vp64, tNE-Vp64+tSBE2-Vp64 and tSBE6-Vp64+tSBE2-Vp64) led to a significant distance increases between Shh and SBE6 (Figure Ch4-6C & Table 9).

Therefore the nuclear distance increase between Shh and SBE6 is a direct effect of *Shh* activation mediated by activators positioned in its regulatory region.

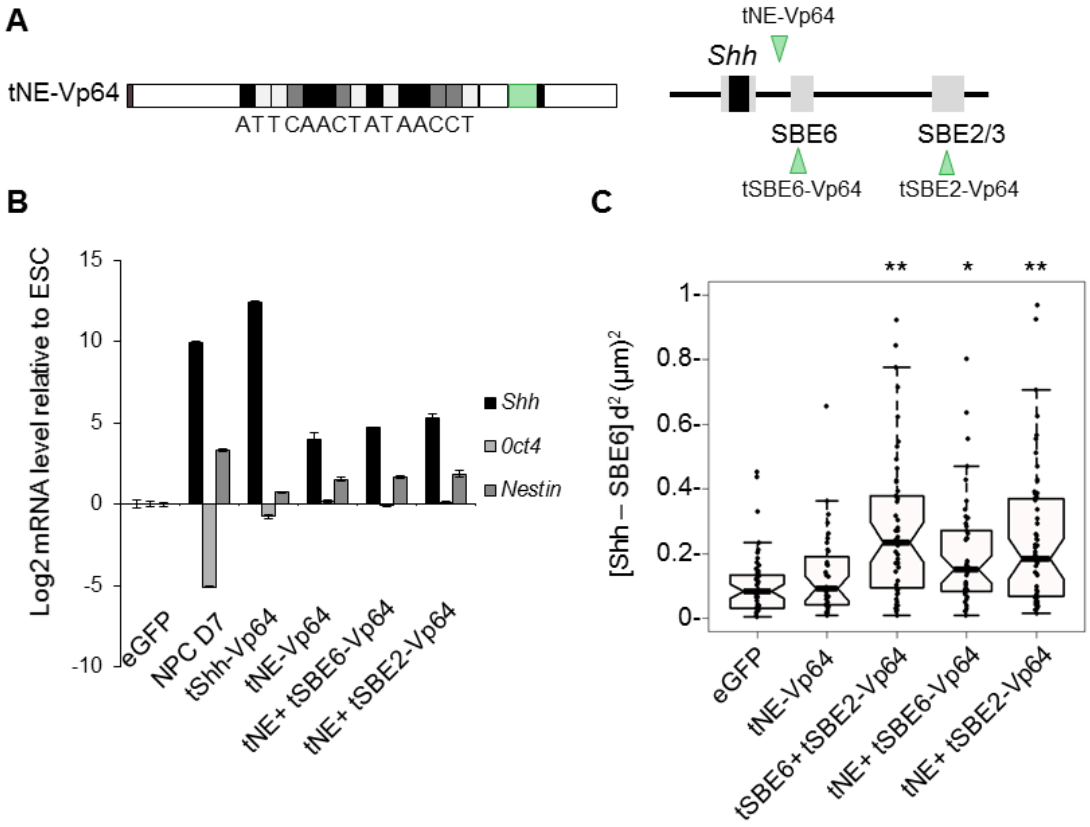

Figure Ch4- 6 Shh Enhancer Bypass in ESC using tNE-Vp64.

A) Schema representing Tale-Vp64 targeting sequence and constructs with map of targeting sites. B) qRT-PCR displaying means (± s.e.m) log2 mRNA levels of *Shh* relative to *Gapdh* after activation in ESC with Tale-Vp64 for three biological replicates normalized to eGFP transfection. C) Boxplots representing Shh-SBE6 squared interprobe distances after Tale-Vp64 transfections.

	Shh-SBE6
tNE-VP64	2.41E-01
tNE+tSBE6-Vp64	5.79E-04
tNE+tSBE2-Vp64	2.82E-02
	7.30E-05

Table 9 P-values from eGFP and Tale-Vp64 transfected ESC Mann-Whitney U tests on FISH distances distribution

Exact p-values of Mann-Whitney U tests significance between ESC and Tale-Vp64 transfected cells revealing the number of replicates.

Linking recruitment of Vp64 and chromatin unfolding, my data led me to think that transcription activator concentration may be important for chromatin unfolding, as it requires VP64 at both SBE6 and SBE2. To test this I recruited Vp128 (twice the amount of Vp64) and could demonstrate that one Vp128 targeted to SBE6 or SBE2 is sufficient to enhance *Shh* activation and to lead to the nuclear distance increase (Figure Ch4-7 & Table 10).

These results indicate that the amount of activator protein matters for the chromatin unfolding to occur, which could happen endogenously by multiple enhancers recruiting more than one activator.

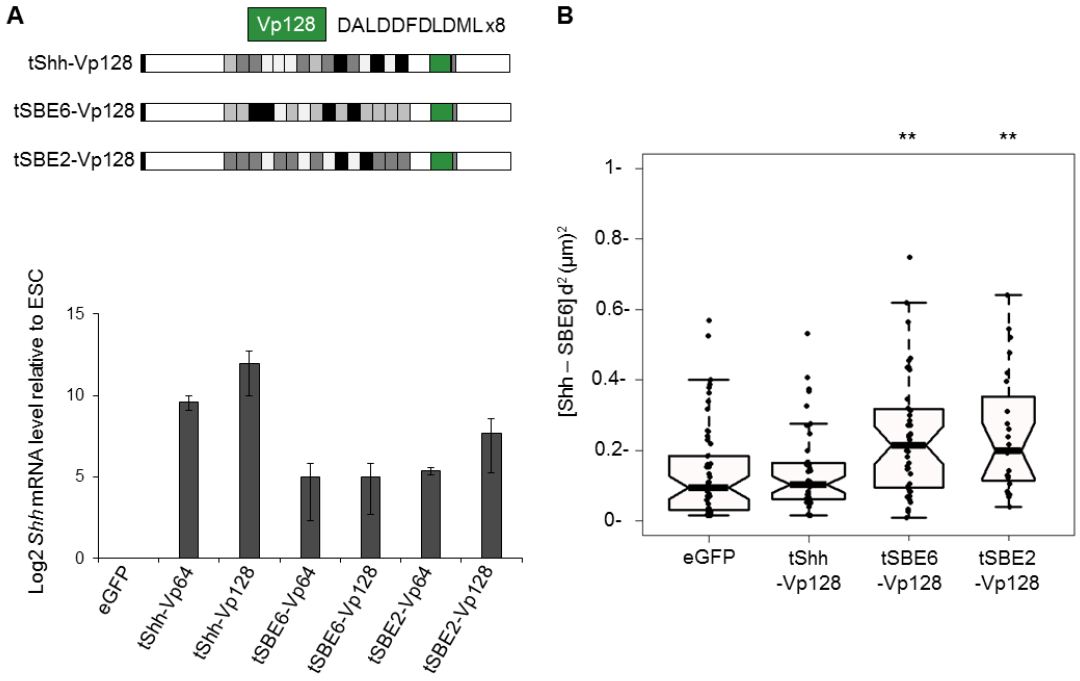

Figure Ch4- 7 Tale-Vp128 is enough and sufficient to activate and unfold the chromatin.

A) Top, Tale-Vp128 constructs targeting the Shh promoter and the SBE6 and SBE2 enhancers. Bottom, qRT-PCR showing mean (± s.e.m) log2 mRNA levels of *Shh* relative to *Gapdh* and normalized to eGFP-transfected cells for Tales-Vp128 and Vp64 for three (Tales-Vp128) and two (Tales-Vp64) biological replicates. B) Boxplots representing Shh-SBE6 squared interprobe distances after FISH on Tale-Vp128 transfected ESC. Asterisks on FISH data represent Mann-Whitney U test significance between Tale-Vp128 and eGFP transfected populations, ** for p-values <0.01.

	Shh-SBE6
tShh-Vp128	7.32E-02
	4.40E-01
tSBE6-Vp128	9.94E-04
tSBE2-Vp128	6.11E-06
	2.17E-03

Table 10 P-values from eGFP and Tale-Vp128 transfected ESC Mann-Whitney U tests on FISH distances distribution

Exact p-values of Mann-Whitney U tests significance between ESC and Tale-Vp128 transfected cells revealing the number of replicates.

4.5 Artificially blocking the chromatin unfolding

In order to have a first idea of the nature of the chromatin unfolding induced by Shh enhancer activation, I next used the binding of CTCF on NE (tNE-CTCF)- a known enhancer insulator (Bell et al., 1999) to see if it could prevent the Shh-SBE6 distance increase and, if so, if that affected *Shh* activation. A ChIP for CTCF confirmed the protein recruitment on NE sequence (Ch4-9A). Interestingly, after a co-transfection of tNE-CTCF together with the Enhancer-Tale-Vp64, Shh-SBE6 squared interprobe distance remained unchanged compared to eGFP transfected ESCs; suggesting that the binding of a Tale-CTCF between Shh and SBE6 could block chromatin unfolding (Figure Ch4-8B & Table 11). This would be consistent with a tracking model in which a change in chromatin is propagated between enhancer and promoter along the length of the chromatin fibre. I was able to reproduce the same effect as observed with tNE-CTCF, using tNE without any fusion (tNE-Delta, Tale without any protein fused) (Figure Ch4-8B & Table 11). However the introduction of tNE-Vp64 alongside tSBE6+tSBE2-Vp64 does not block the chromatin unfolding, indicating a specific action of blocking agent from tNE-CTCF or tNE-Delta (Figure Ch4-9B). These results indicate that the chromatin unfolding can be blocked if a protein bound in between is not an activator as tNE-Vp64 did not disturbed the chromatin unfolding.

Interestingly the binding of tNE-CTCF or tNE-Delta between SBE6 and Shh significantly decreased *Shh* activation but did not abolish it (Figure Ch4-8C). Single cell qRT-PCR showed that some cells still expressed *Shh*. Analysis of the population distribution on *Shh* mRNA levels revealed that this is due to a heterogeneous distribution of the population (Figure Ch4-8D). This indicates that *Shh* expression levels may be dependent on the number of plasmid each cells received, even if the majority will still have the three plasmids and neither show a distance increase nor *Shh* mRNA production.

Together these results show that interfering with the chromatin unfolding upon co-activation of SBE6.1 and SBE2 via adding a protein complex such as Tale-Delta or Tale-CTCF in between, does not lead to a proper *Shh* activation. Indicating perhaps that the brain-enhancer communication model might follow a tracking-like mechanism.

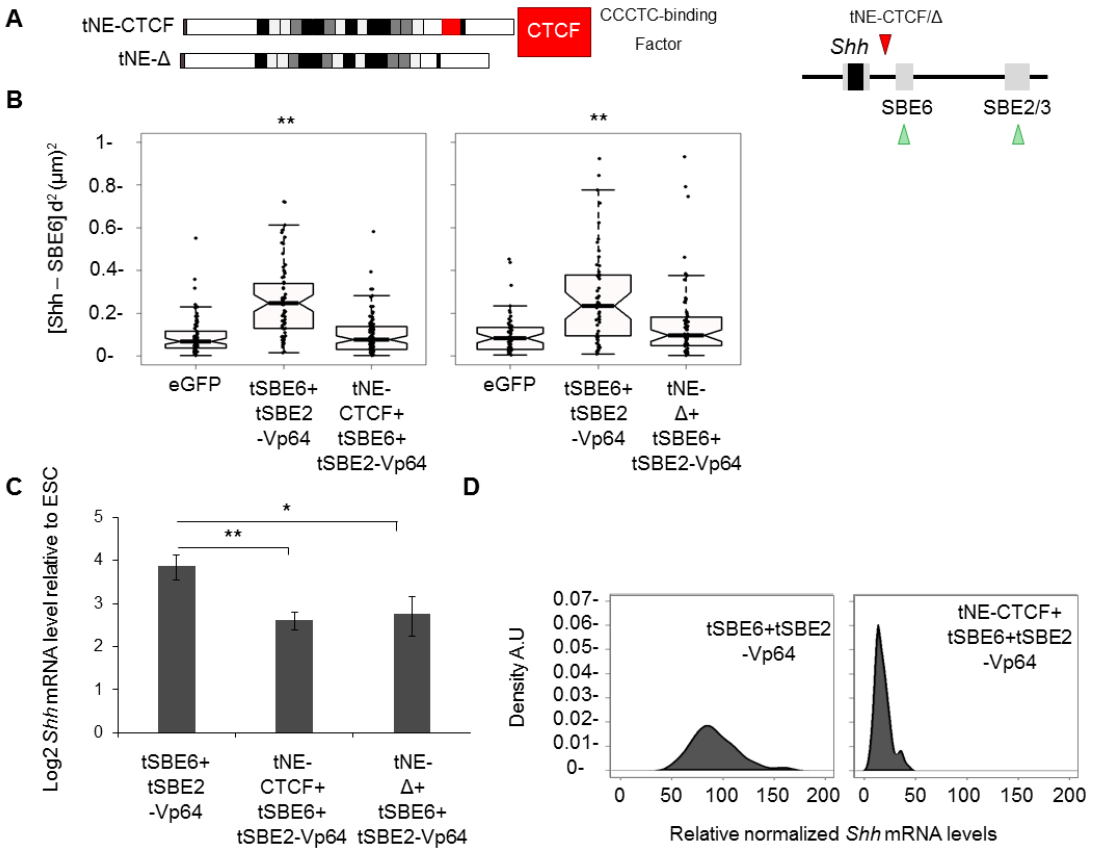

Figure Ch4- 8 Artificially blocking the chromatin unfolding.

A) tNE-CTCF and tNE-Delta (Δ) constructs with map of the Tales targeting. B) Boxplots representing Shh-SBE6 squared interprobe distances upon co-transfection with activators and blocker. C) qRT-PCR displaying means (± s.e.m) log2 mRNA levels of *Shh* relative to *Gapdh* after co-transfection with activators and blocker normalized to eGFP transfection for four (tNe-CTCF) and three (tNE-Δ) biological replicates. Asterisks represent p-values for one-tailed Student t-test between conditions. ** for p-values <0.01 and * for p-values <0.05. D) Kernel density plots showing *Shh* mRNA expression of single tSBE6+tSBE2-Vp64 cells with and without tNE-CTCF relative to *Gapdh* and normalised to expression in eGFP transfected ESC.

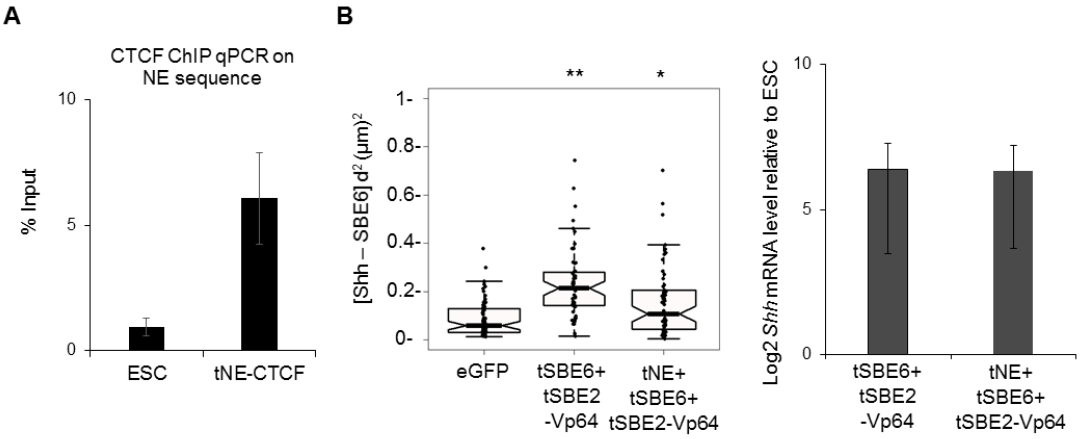

Figure Ch4- 9 Vp64 in between Shh and Shh-Brain-Enhancer does not perturb the unfolding.

A) CTCF ChIP qPCR on NE sequence after tNE-CTCF transfection. qPCR has been normalised to input and Actin sequence for one biological replicate. B) Left panel shows boxplots representing Shh-SBE6 squared interprobe distances after FISH on Tale-Vp64 transfected ESC with and without tNE-Vp64. Asterisks on FISH data represent Mann-Whitney U test significance between Tale-Vp64 and eGFP populations, ** for p-values <0.01. Right, qRT-PCR showing means (± s.e.m) log2 mRNA levels of *Shh* relative to *Gapdh* and normalized to eGFP transfected cells for Tales-Vp64 for two biological replicates.

	Shh-SBE6
tNE-Δ+tSBE6+tSBE2-Vp64	1.68E-01
	5.96E-01
tNE-CTCF+tSBE6+tSBE2-Vp64	6.26E-01
	6.02E-01
tNE+tSBE6+tSBE2-Vp64	5.02E-03
tNE-CTCF+tSBE2-Vp128	4.56E-01

Table 11 P-values from eGFP and Tale-Vp64/Tale-Δ/Tale-CTCF transfected ESC Mann-Whitney U tests on FISH distances distribution

Exact p-values of Mann-Whitney U tests significance between ESC and Tale transfected ESCs revealing the number of replicates.

4.6 5C Chromatin conformation capture carbon-copy (5C) on ESC with Tale-Vp64

As in the previous chapter (Chapter 3, Section 3.7), I wanted to see if 5C could reveal, and perhaps confirm, the chromatin unfolding induced at the Shh regulatory domain by Vp64s recruitment.

5C heat-maps of Tale-Vp64 transfected ESC confirmed the observation of a Shh TAD closed by a loop formed with sequence spanning the ZRS and Shh both in cells expressing and non-expressing *Shh* (Williamson et al., 2016) (Figure Ch3-7A, Ch3-8A, Ch4-10A and Ch4-11A).

With the first degree of zooming in the Shh regulatory region, the 5C heat-maps also confirm that the sequence spanning Shh interacts with the ZRS independently of the expressing state of *Shh* and the chromatin unfolding (the Shh-SBE6 distance observed by FISH) (Figure Ch4-10B and Ch4-11B).

Deeper zoom and comparison between ESC and NPC heat-maps (Figure Ch4-10C) and various Tale-Vp64 transfected ESC 5C heat-maps did not reveal the loss of at the CTCF interaction as shown in NPC (Figure Ch3-7C, Ch3-8C, Ch4-10C and Ch4-11C). Therefore the CTCF loss of interaction is not the principal component of the chromatin unfolding as this interaction remains in tSBE6-Vp64+tSBE2-Vp64 (Figure Ch4-10C).

However, the broad loss of interaction in the region between Shh and SBE6 pointed by an arrowhead seems to be consistent in all replicates undergoing chromatin unfolding (Figure Ch3-7C & Ch4-10C). Moreover, these 5C heat-maps support our view that *Shh* is not interacting with a missed enhancer or any other sequence upon activation in NPC or Tale-Vp64s (Figure Ch3-7, Ch3-8, Ch4-10 and Ch4-11).

However, as previously concluded, 5C heat-maps in that case are more qualitative than quantitative, therefore I can only use them to test that no other chromatin conformation is occurring in that locus that contradict our view.

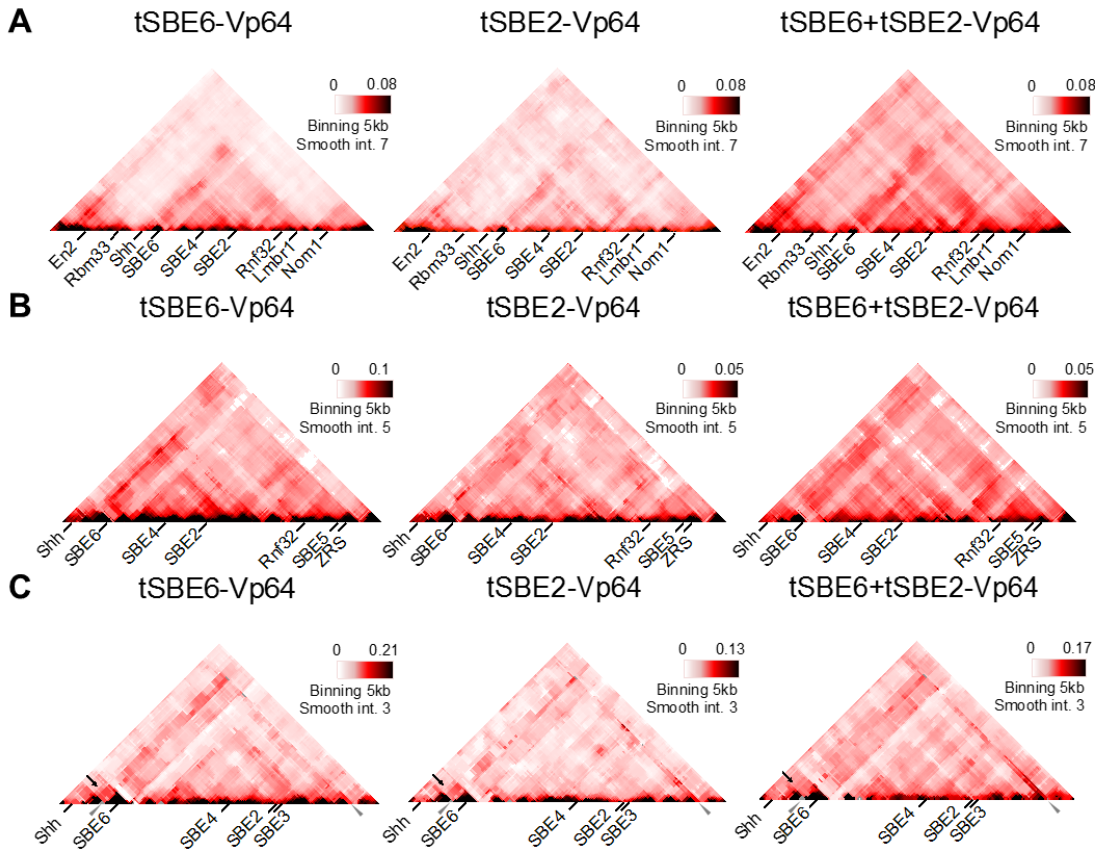

Figure Ch4- 10 Chromatin conformation capture carbon-copy (5C) heat-maps of the Shh regulatory region in ESC after Tale-Vp64 transfection.

A) 5C heat-maps beyond Shh regulatory region (chr5:28317087-30005000) with 5kb binning and smoothing with the median of 7 surrounding interactions frequencies. B) 5C heat-maps of the Shh regulatory region (chr5:28750000-29750000) with 5kb binning and smoothing with the median of 5 surrounding interactions frequencies. C) 5C heat-maps of the Shh regulatory region (chr5:28750000-29450000) with 5kb binning and smoothing with the median of 3 surrounding interactions frequencies of a biological replicate. Black arrows highlights the differences with ESCs observed in Figure Ch3-7 and grey arrowhead show CTCF sites.

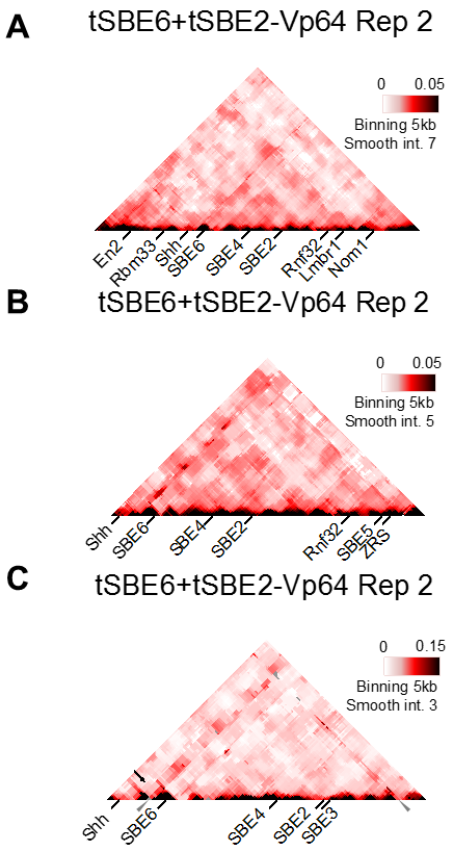

Figure Ch4- 11 Chromatin conformation capture carbon-copy (5C) heatmaps of the Shh regulatory region in tSBE6+tSBE2-Vp64 biological replicate.

A) 5C heatmaps beyond Shh regulatory region (chr5:28317087-30005000) with 5kb binning and smoothing with the median of 7 surrounding interactions frequencies. B) 5C heatmaps of the Shh regulatory region (chr5:28750000-29750000) with 5kb binning and smoothing with the median of 5 surrounding interactions frequencies. C) 5C heatmaps of the Shh regulatory region (chr5:28750000-29450000) with 5kb binning and smoothing with the median of 3 surrounding interactions frequencies of a biological replicate. Black arrows highlights the differences with ESCs observed in Figure Ch3-7 and grey arrowhead show CTCF sites.

4.7 Conclusions

Together these data suggest that the *Shh* activation that occurs upon its brain-enhancer activation from SBE6 and SBE2, promotes a large chromatin unfolding of a 100kb domain that can be blocked by recruiting a large protein complex such as Tale-Delta or Tale-CTCF, and that this is linked with enhancement of *Shh* expression.

Thus, *Shh* neural enhancer-promoter communication seems to be mediated by a spreading or tracking of an activation signal through the intervening chromatin region that induces chromatin unfolding and more mRNA production.

Chapter 5: Analysing the mechanism of the chromatin unfolding

5.1 Which Factor is responsible for the chromatin unfolding?

5.1.1 Supercoiling, a physical twist leading to chromatin unfolding?

DNA supercoiling occurs endogenously as is an intrinsic consequence of DNA distortion. It has been previously described as critical for regulating gene expression and DNA replication (Dorman, 1991). DNA supercoiling can change chromatin fibre conformation and inter-probe distances measured by FISH (Naughton et al., 2013). I therefore tested whether the chromatin folding I have detected in the previous two Chapters might be due to negative supercoiling (underwound DNA structure). Bleomycin is an antibiotic that introduces DNA strand breaks and that therefore releases supercoiling (Naughton et al., 2013).

Although this data has been repeated only once, it seems that releasing the torsional stress with bleomycin does not relax chromatin structure in the Shh-SBE6 region in ESC nor in NPC (Figure Ch5-1). There is no significant difference between NPC and treated NPC, as nicking DNA did not release any distortion, and the chromatin unfolding is still significant upon treatment in NPC.

However, without being significant, the data between ESC and ESC treated could indicate that bleomycin treatment might still have an effect on FISH inter-probe distances as the distribution becomes slightly more spread. This could be due to a perhaps overwound DNA structure of this loci in ESCs that is lost during differentiation participating in the Shh-SBE6 chromatin unfolding without being the leading mechanism.

Therefore this preliminary result indicates that DNA supercoiling is not the primary determinant of the chromatin structure rearrangement upon neural differentiation.

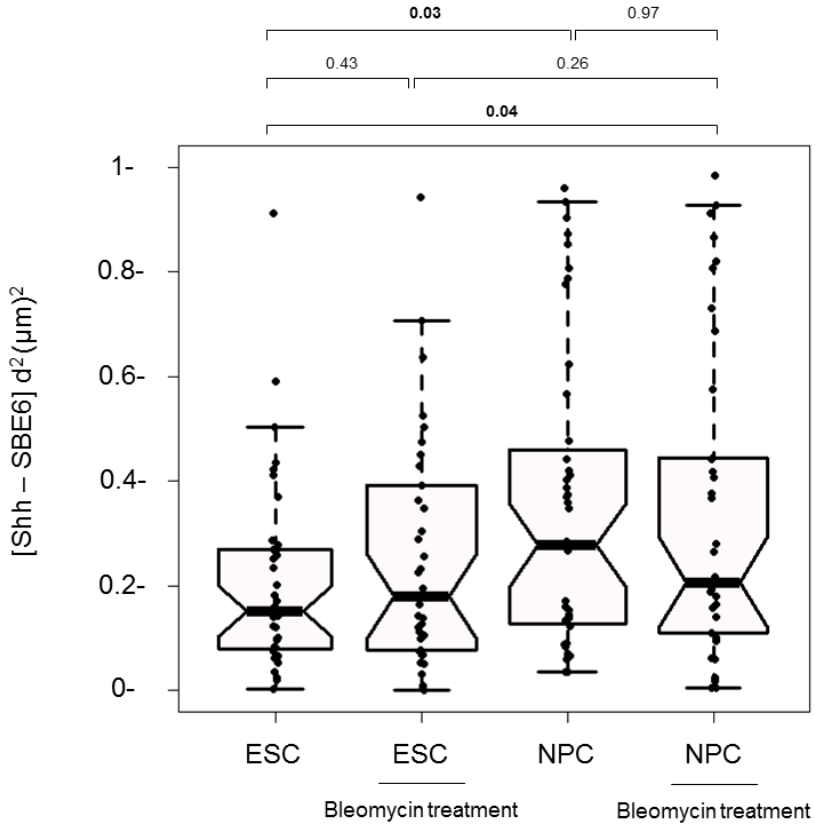

Figure Ch5- 1 Effect of Supercoiling on the NPC chromatin unfolding.

Boxplots representing Shh-SBE6 squared interprobe distances in ESC and NPC treated or not treated with bleomycin that releases supercoiling. N=1.

5.1.2 Recruiting endogenous transcriptional co-activators, Mediator and p300

As Vp16 is an ectopic (viral) transcriptional activator, I next wanted to determine if recruiting endogenous transcriptional activators and co-activators could reproduce my results.

Vp16 is known to interact with the Mediator complex, which is a known co-activator of transcription (Allen and Taatjes, 2015; Yin et al., 2014). Mediator has several binding partners such as RNA polymerase II (Kim et al., 1994), Vp16 (Vojnic et al., 2011), and transcription factors (Taatjes, 2010). By ChIP, it is found at active enhancer and promoters and it might be involve in chromatin structure as it can work alongside cohesin to rearrange chromatin upon enhancer driven gene activation (Kagey et al., 2010; Visel et al., 2009).

I focused on Med25, a subunit of the tail domain of Mediator that has the ability to bind Vp16, and therefore which could be recruited upon Tale-Vp64 transfection. I fused my Tales with the N-terminal Von Willebrand factor A domain (VWA domain) of Med25 that is responsible for Mediator binding and thus could recruit the whole Mediator complex (Figure Ch5-2A & Ch5-3A). I showed that using tShh-VWA, tSBE6-VWA and tSBE2-VWA, I could reproduce all my previous data generated with Tale-Vp64 (Figure Ch5-2B & Ch5-3B and Table 12).

However, Med25-VWA recruitment induces very low levels of mRNA expression for *Shh* but also *Ptn* and *Nrp1* using Tales targeting their promoter (Therizols et al., 2014) (Figure Ch5-3C). Indeed, mRNA levels are considerably inconsistent and reach similar amplitude than mRNA levels after a Tale-Δ transfection (Figure Ch5-3D). This could indicate that the role in this context might principally be to remodel chromatin rather than to activate expression *per se.*

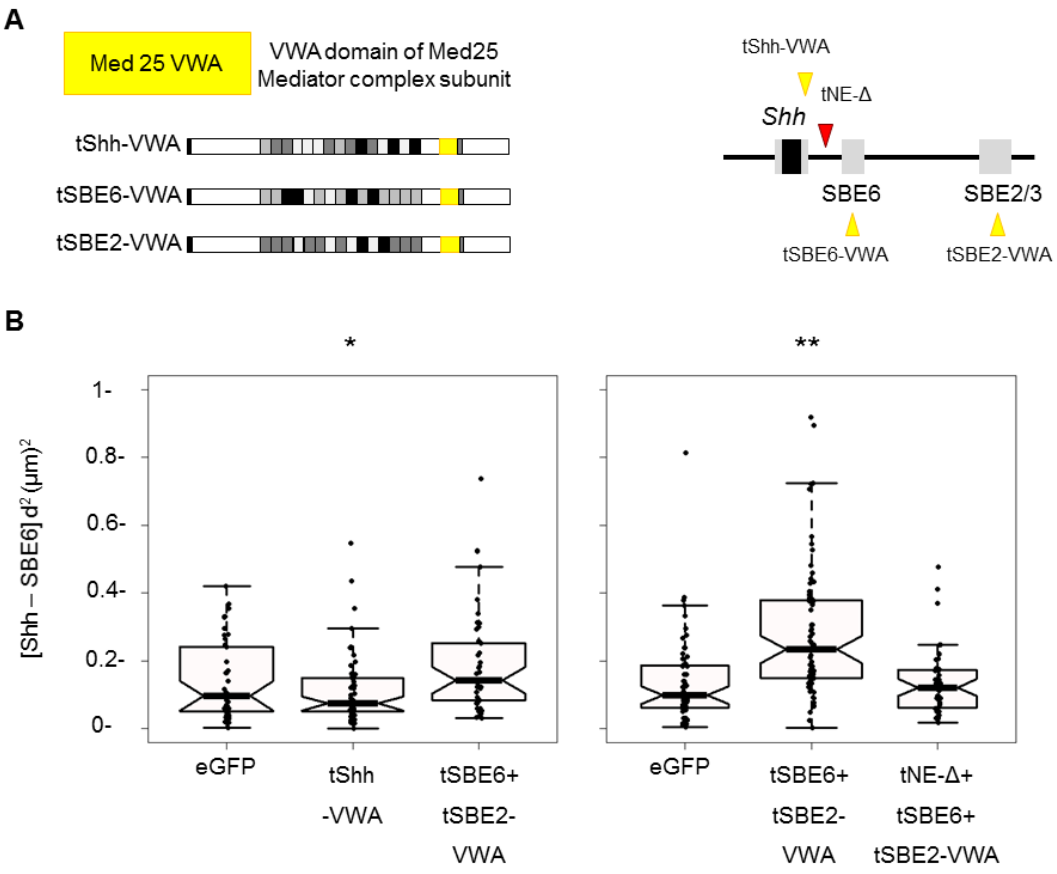

Figure Ch5- 2 Recruitment of the VWA domain of Med25, a subunit of Mediator complex.

A) Tale-VWA constructs and targeting. B) Boxplots representing Shh-SBE6 squared interprobe distances upon co-transfection with Tale-VWA and blocker.

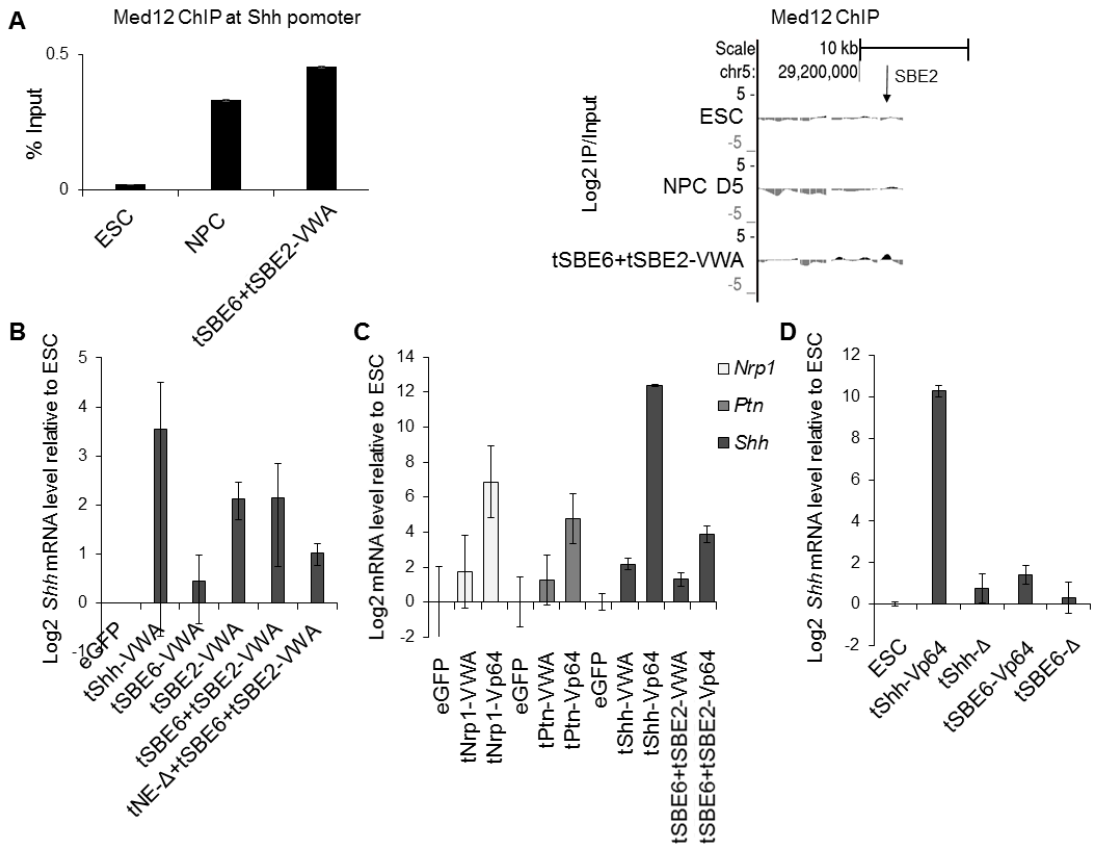

Figure Ch5- 3 Mediator Complex recruitment.

A) Left- Med12 ChIP qPCR on Shh promoter sequence in NPC or after tSBE6+tSBE2-VWA transfection. Right- Med12 ChIP-chip in NPC or after tSBE6+tSBE2-VWA transfection view on SBE2 sequence. B) qRT-PCR displaying means (± s.e.m) log2 mRNA levels of *Shh* relative to *Gapdh* after co-transfection with Tale-VWA normalized to eGFP transfection for four biological replicates. C) qRT-PCR showing means (±s.e.m) log2 mRNA levels of *Shh*, *Ptn* and *Nrp1* relative to *Gapdh* and normalized to eGFP transfected cells for Tales-Vp64 and Tales-VWA for one biological replicate. D) qRT-PCR showing means (±s.e.m) log2 mRNA levels of *Shh* relative to *Gapdh* and normalized to eGFP transfected cells for Tales-Vp64 and Tales-Δ for one biological replicate.

The p300, a histone acetyl-transferase (HAT), is known to catalyse acetylation of lysine 27 of histone H3 (H3K27ac) and strongly occupies active enhancer (Visel et al. 2009). Moreover, it has been recently demonstrated that recruiting p300 to enhancers using the dCas9 from the CRISPR system was enough to trigger gene activation of their cognate promoter (Hilton et al., 2015). Cas9 is an RNA-guided DNA endonuclease from a type II CRISPR system, a bacterial immune system that directs the degradation of foreign nucleic acids. dCas9 is a dead version of the endonuclease. Small guide RNAs (sgRNA) could then guide dCas9 lacking endonuclease activity to desired DNA sequences (Gilbert et al., 2013; Qi et al., 2013). Like Tales-Vp64, dDCas9-Vp160 (ten repeats of Vp16) had also been used to activate endogenous sequences (Cheng et al., 2013). Recently, the catalytic core domain of p300 has been fused to a catalytically dead Cas9 (dCas9–p300Core), dCas9–p300Core, and was able to activate transcription when targeted to promoters and even when targeted to enhancers (Hilton et al., 2015). Thus p300 may be able to activate gene expression through catalysing H3K27ac marks at proximal and distal regulatory sequences.

I used the dCas9-p300Core from Hilton et al., 2015 to compare with the levels of activation obtained by Tales-Vp64 in the *Shh* context. I used one guide RNA on SBE6 and one on SBE2 with dCas9-p300Core (cSBE6-p300 and cSBE2-p300).

Preliminary results indicate that cShh-p300, cSBE6-p300 and cSBE2-p300 fail to lead to *Shh* activation but cSBE6+cSBE2-p300 transfection still led to the nuclear distance increase between Shh and SBE6 (Figure Ch5-4 & Table 12).

This result strongly supports my data previously discussed in Chapter 4, that is to say that a double activation mark is needed to promote the chromatin unfolding. As p300 is already a known marker of active enhancers I did not pursue the investigation of p300 targeting.

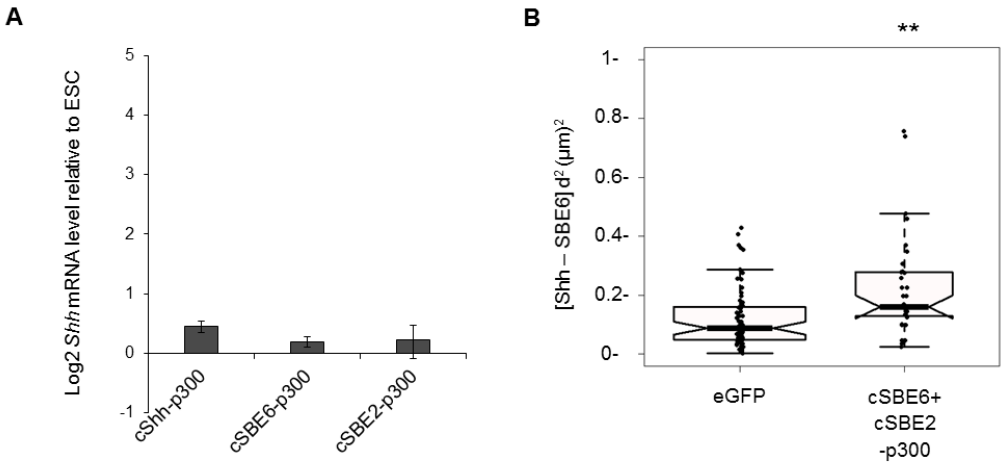

Figure Ch5- 4 Recruitment of p300Core via dCas9 and one gRNA on SBE6 and SBE2.

A) qRT-PCR displaying means (± SEM) log2 mRNA levels of Shh relative to Gapdh after co-transfection with p300 normalized to eGFP transfection for three biological replicates. B) Boxplots representing Shh-SBE6 squared interprobe distances upon co-transfection with dCas9-p300 and SBE6 and SBE2 guide RNA.

	Shh-SBE6
tNE-VWA	4.89E-01
tSBE6-VWA	3.59E-01
	1.98E-01
tSBE6+tSBE2-VWA	6.16E-08
	3.96E-02
tNE+tSBE6-VWA	3.49E-01
tNE+tSBE2-VWA	2.12E-02
tNE-Δ+tSBE6+tSBE2-VWA	6.27E-01
cSBE6+cSBE2-p300	4.36E-04

Table 12 P-values from eGFP and Tale-VWA or dCas9-p300 transfected ESC Mann-Whitney U tests on FISH distances distribution

Exact p-values of Mann-Whitney U tests significance between ESC and Tale-VWA or dCas9-p300 revealing the number of replicates.

5.1.3 Recruiting an endogenous transcription factor: SIX homeobox 3 (SIX3) recruitment to SBE2 leads to chromatin unfolding without activating *Shh*

Point mutation of a SIX homeobox 3 (SIX3) transcription factor binding domain in SBE2 or a mutation of SIX3 that reduced its binding affinity have been shown to affect *Shh* expression in the forebrain and to lead to severe cases of Holoprosencephaly (HPE) (Geng et al., 2008; Jeong et al., 2008). Therefore there is a direct link between SIX3 binding at SBE2 and *Shh* expression in the forebrain. I therefore investigated whether SIX3 binding could be sufficient to activate *Shh* using Tale-based recruitment to *Shh* promoter, SBE6 or SBE2 (Figure Ch5-5A & Ch5-5B).

The recruitment of Tale-SIX3 is not able to lead to consistent *Shh* activation, but leads to the chromatin unfolding only when bound to SBE2. (Figure Ch5-5C & Table 13). Furthermore, as previously shown, introducing tNE-CTCF was enough to prevent the chromatin to unfold when co-transfected with tSBE2-SIX3.

These results reveal a crucial role for SIX3 binding on SBE2, as impairing its binding would probably prevent chromatin unfolding in the natural context and thus would not be able to enhance *Shh* expression (Figure Ch5-5D).

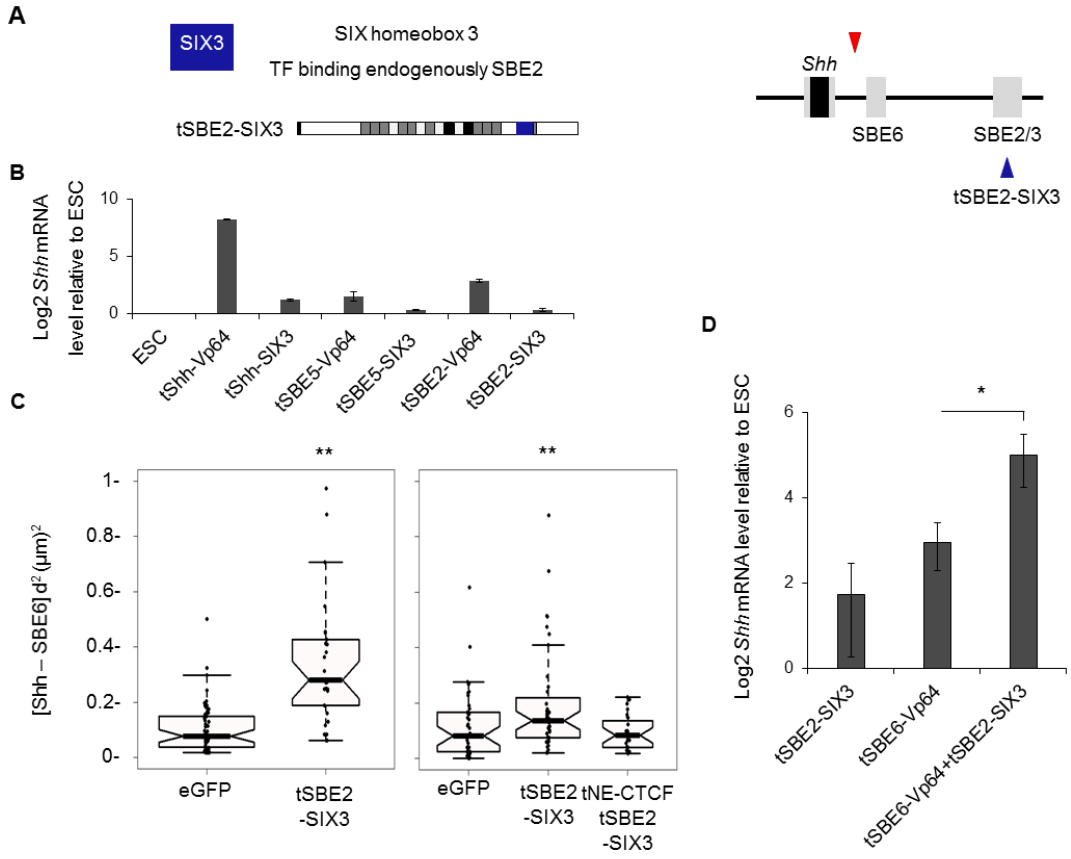

Figure Ch5- 5 SIX3 recruitment allows chromatin unfolding and enhanced *Shh* expression when activators are present.

A) tSBE2-SIX3 construct and targeting map. B) qRT-PCR showing means (± s.e.m) of log2 mRNA levels of *Shh* relative to *Gapdh* after Tale-Vp64 or Tale-SIX3 and both transfections in ESC in two biological replicates normalized to eGFP. C) Boxplots representing Shh-SBE6 squared interprobe distances after tSBE2-SIX3 and tNE-CTCF transfections. Asterisks on FISH data represent Mann-Whitney U test significance between tSBE2-SIX3 and eGFP populations, ** for p-values <0.01. D) qRT-PCR showing medians (± s.e.m) of log2 mRNA levels of *Shh* relative to *Gapdh* after tSBE2-SIX3, tSBE6-Vp64 and both transfections in ESC in thirteen (tSBE2-SIX3), eleven (tSBE6-Vp64) and nine (tSBE2-SIX3+tSBE6-Vp64) biological replicates normalized to eGFP. Asterisks represent p-values for one-tailed Student t-test between conditions. * for p-values <0.05.

	Shh-SBE6
tShh-SIX3	5.94E-01
tSBE6-SIX3	3.34E-01
tSBE2-SIX3	1.09E-08
	1.98E-05
	1.18E-02
tNE-CTCF+tSBE2-SIX3	7.68E-01

Table 13 P-values from eGFP and Tale-SIX3 transfected ESC Mann-Whitney U tests on FISH distances distribution

Exact p-values of Mann-Whitney U tests significance between ESC and Tale-SIX3 revealing the number of replicates.

5.2 Investigating chromatin marks that might contribute to chromatin unfolding

Overall my results highlight a visible level of chromatin decompaction over a 100kb region 5' of Shh that occurs upon brain-enhancer activation of *Shh* during neural differentiation and upon ectopic recruitment of activators and co-activators to the gene desert upstream of Shh. As the chromatin unfolding can be blocked with an introduction of Tale-Delta or Tale-CTCF in the intervening region, my data is consistent with a long-range tracking mechanism and a spreading of altered chromatin structure.

5.2.1 Spreading of H3K27ac?

As the unidirectional spreading of the CBP/p300 histone acetyl-transferases and the histone H3 acetylation have been previously involved with promoter activation via tracking mechanisms from their enhancer recruitment (Gribnau et al., 2000; Hatzis and Talianidis, 2002; Kim and Dean, 2004; Masternak et al., 2003; Spicuglia et al., 2002; Wang et al., 2005; Zhao and Dean, 2004), I first investigated H3K27ac tracking in our system.

I performed a time lapse after co-transfection of Tale-Vp64 targeted to Shh-brain enhancer to monitor the increase of nuclear distance. It appears that only 18h after Tale-Vp64 transfection, the chromatin unfolding becomes significant (Figure Ch5-6). I then performed H3K27ac ChIP 18h after transfection of Tale-Vp64s.

H3K27ac Chip-chip revealed no spreading of acetylation marks in NPC and no tendency of spreading in Tale-Vp64 transfected cells (Figure Ch5-7). On the contrary, the acetylation marks are precisely deposited on the active enhancer and *Shh* promoter or gene body. We can also note the discrete acetylation on the site where Tale-Vp64 binds, certainly due to the recruitment of p300 via Vp16 (Kundu et al., 2000).

Therefore p300 does not appear to be involved in the spreading of the signal of the chromatin unfolding.

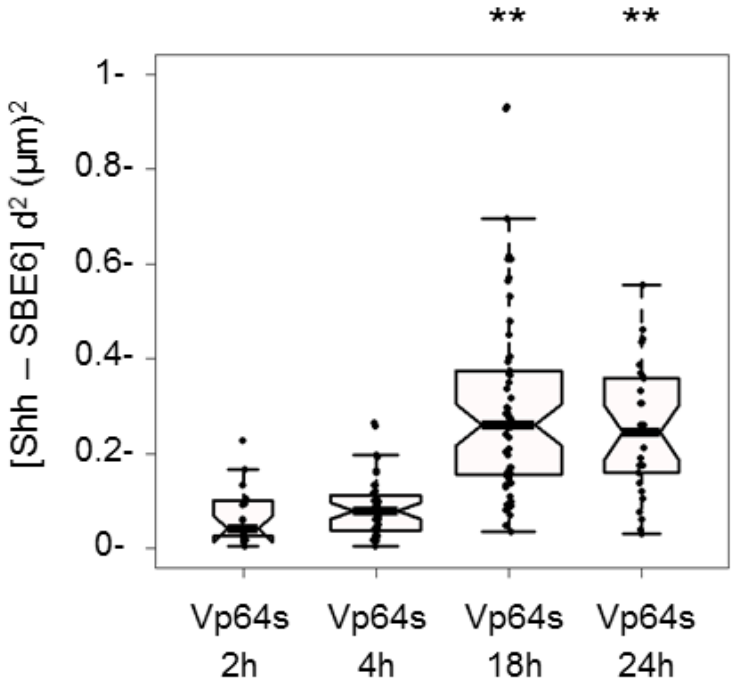

Figure Ch5- 6 Chromatin unfolding becomes specific 18h after transfections

Boxplots representing Shh-SBE6 squared interprobe distances after tSBE6+tSBE2-Vp64 2h, 4h, 18h and 24h after transfections. Asterisks on FISH data represent Mann-Whitney U test significance between Tale-Vp64 18h or 24h compared to 2h after transfection, ** for p-values <0.01.

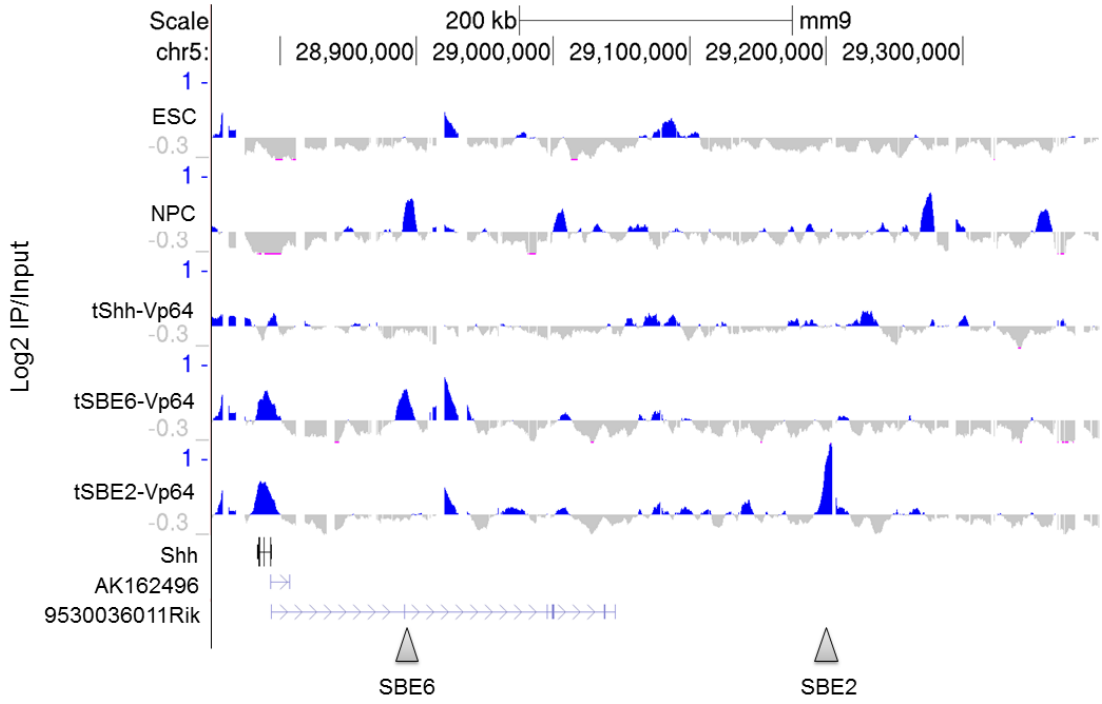

Figure Ch5- 7 H3K27ac does not spread along the Shh Regulatory region.

UCSC Genome browser view of chr5:28,750,000-29,400,000 build mm9 displaying log2 H3K27ac ChIP relative to input chromatin on Shh regulatory region on ESCs and NPCs and after Tales-Vp64 transfection in ESC, one biological replicate. Arrowheads represent enhancer location SBE6, SBE2.

5.2.2 Spreading of RNA Polymerase II?

RNA polymerase II and TBP tracking on the chromatin fibre has been previously established between the HS2 enhancer of the LCR and the epsilon-globin promoter (Zhu et al., 2007). Indeed, RNA polymerase II can track along the intervening DNA and induce the synthesis of short, polyadenylated RNAs.

I isolated RNA from the nuclei of ESC and NPCs to prevent the dilution of the potential non-coding intervening transcripts by cytoplasmic mRNA and the degradation by the exosome. I then assayed the distribution of these transcripts using microarrays that tile the entire *Shh* regulatory region.

However comparison of eGFP-transfected cells with NPC and other Tale-Vp64 transfected cells did not reveal any evidence of tracking of RNA pol II by hybridising RNA on chip (Figure Ch5-8).

Together these analyses indicate that some other kind of protein complex is involved in the spreading of the chromatin-unfolding signal

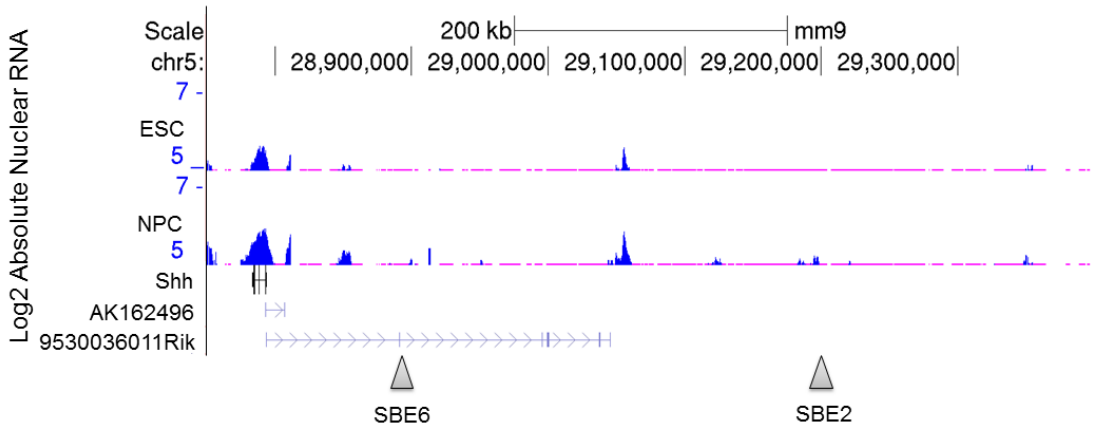

Figure Ch5- 8 No nascent transcripts emerge from the SBEs to Shh.

UCSC Genome browser view of chr5:28,750,000-29,400,000 build mm9 displaying log2 absolute nuclear RNA on chip for ESC and NPC on 3 biological replicates combined. Arrowheads represent enhancer location SBE6, SBE2.

5.2.3 Decompaction via removal of Polycomb marks?

Previous work in the Bickmore lab has demonstrated that in absence of PRC1 and PRC2 Polycomb components, there is visible chromatin decompaction chromatin at Polycomb target genes, and that this is due to a non-catalytic function of PRC1 (Eskeland et al. 2010). Polycomb-mediated repression has been extensively linked with methylation of the Lys 27 of histone H3 (H3K27me3) (Cao and Zhang, 2004).

Shh is a known Polycomb target gene in ESC as revealed by ENCODE ChIP-seq data (Hawkins et al., 2010)

Performing Chromatin ImmunoPrecipitation on arrays tiling the whole *Shh* region (ChIP-on-chip), for the Polycomb mark H3K27me3, and for H3K27ac associated with active chromatin revealed that interestingly *Shh,* when active, does not significantly lose its H3K27me3 marks (Figure Ch7-9). However activation of *Shh* via Vp64 tethered at SBE6.1 or SBE2 leads to a more extensive domain of H3K27ac that spreads across *Shh* and the wider genomic region (Figure Ch5-7) indicating that there must be some loss of H3K27me3.

To further complement that view, I used a Ring1b$^{-/-}$ cell line that is deficient in Polycomb machinery and that has been used to link Ring1b knock out with Polycomb loci decompaction (Eskeland et al., 2010b).

FISH on wild type (WT) and Ring1b$^{-/-}$ ESCs transfected or not with Tale-Vp64s showed that co-transfection of the tSBE6-Vp64+tSBE2-Vp64 in Ring1b$^{-/-}$ ESCs does not lead to a significant decompaction compared to Ring1b$^{-/-}$ ESCs. However, the decompaction between WT and Ring1b$^{-/-}$ ESC is not as strong as the one that occurs upon Shh-Brain-Enhancer activation with Tale-Vp64s. Consistently with this data, FISH inter-probe distances measured in Ring1b$^{-/-}$ ESC are significantly smaller than distances measured after Tale-Vp64s transfection in WT ESCs, affirming that the nuclear distance increase observed in Ring1b$^{-/-}$ ESC is not comparable in term of amplitude with the once occurring in neural differentiation (Figure Ch5-10). Furthermore ESC Ring1b$^{-/-}$ transfected with the cocktail of Tale-Vp64 could show a distance increase compared to the ESC WT control (Figure Ch5-10).

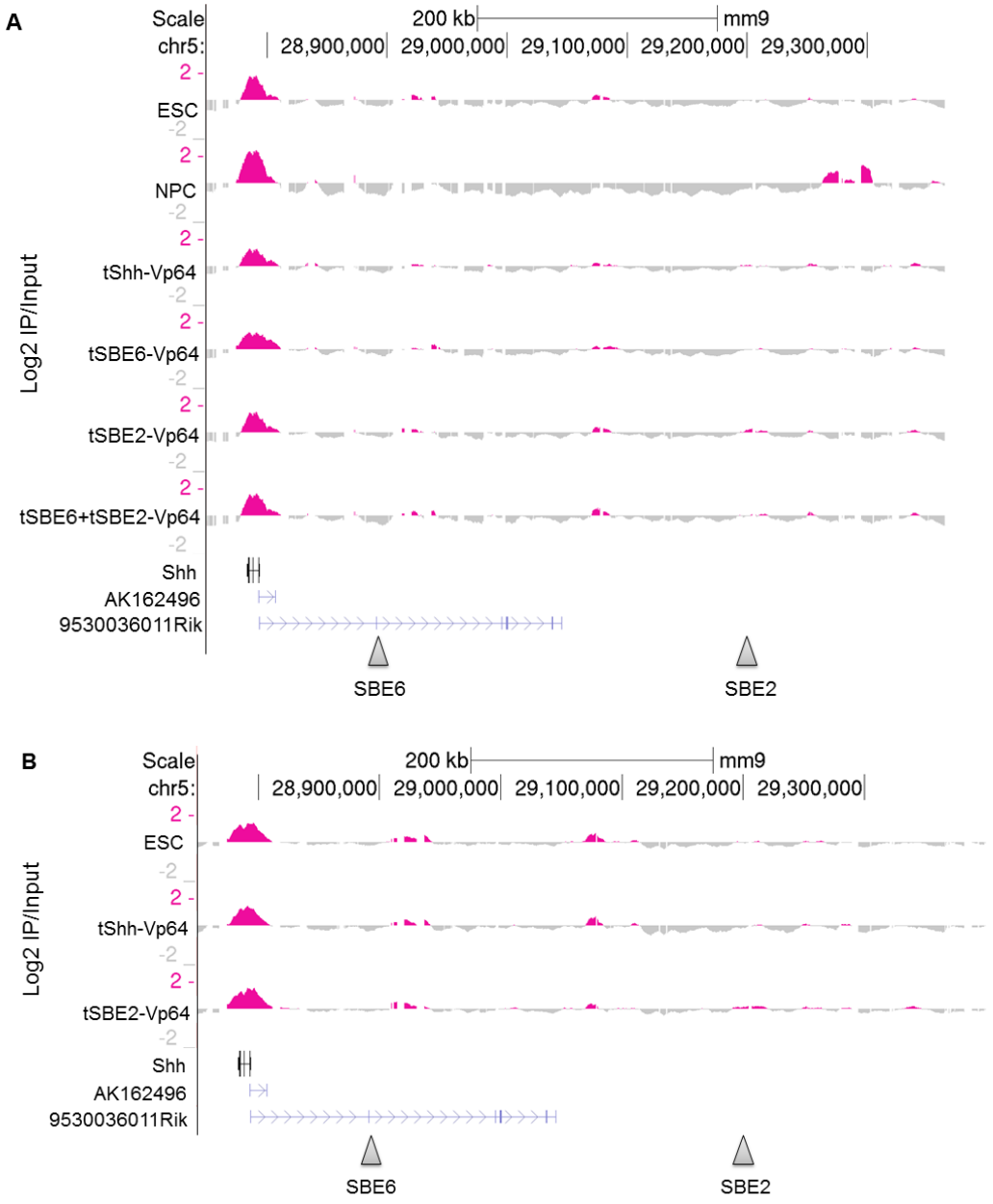

Figure Ch5- 9 Polycomb marks do not disappear upon Shh activation during neural differentiation or Tale-Vp64 activation.

UCSC Genome browser view of chr5:28,750,000-29,400,000 build mm9 displaying log2 H3K27me3 ChIP relative to input on Shh regulatory region after Tales-Vp64

transfection in ESC or in NPC, two biological replicate displayed A and B. Arrowheads represent enhancer location SBE6, SBE2.

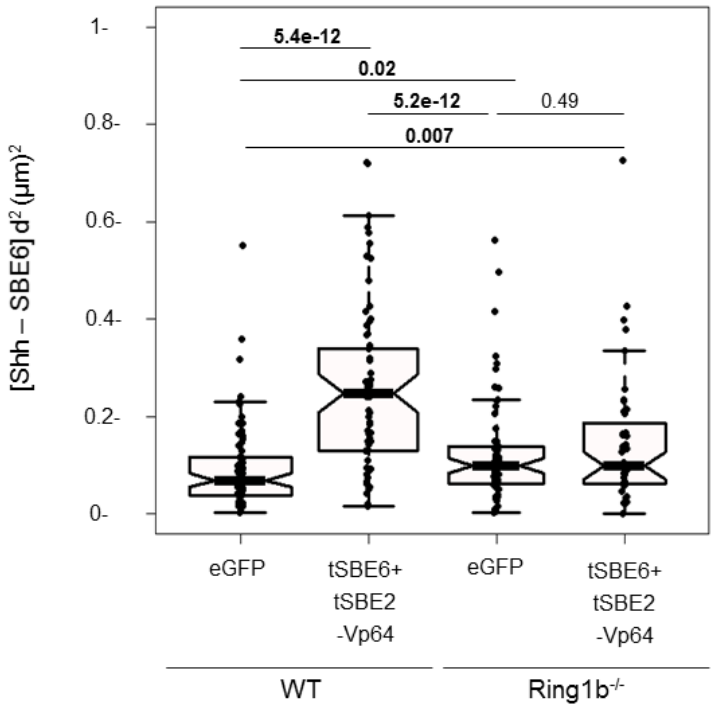

Figure Ch5- 10 Absence of Polycomb marks does not explain the chromatin unfolding.

Boxplots representing Shh-SBE6 squared interprobe distances after Tale-Vp64 transfections in wild-type ESC (WT) or in Ring1b-/- ESC. Asterisks on FISH data represent Mann-Whitney U test significance between populations.

As for the bleomycin experiment, this experiment indicates that the loss of Polycomb marks on the Shh locus might take part in the chromatin unfolding between Shh and SBE6, but is not the primary determinant. Together these data point toward the fact that Polycomb machinery solely cannot explain the chromatin unfolding that occurs upon Shh-brain-enhancer activation.

5.2.4 Poly (ADP-ribose) polymerase-1 (PARP1)

Given that the chromatin decompaction I detected does not seem due to loss of Polycomb, I looked for other chromatin-modifying systems that have been associated with large-scale chromatin decompaction.

PARP1 has been shown to be involved in chromatin remodelling and gene expression (Tulin A et al., 2003, Ogino H et al., 2007). Indeed PARP1 can act on chromatin structure resulting in changes in gene expression using different mechanisms. It alters chromatin structure by competing with histone H1 and nucleosome binding (Kim et al., 2004; Krishnakumar and Kraus, 2010; Krishnakumar et al., 2008) or by acting as a chromatin remodeller and PARylates core histones, histone H1 or other chromatin architectural proteins (Gottschalk et al., 2009; Petesch and Lis, 2008; Timinszky et al., 2009). PARP1 is also associated with active histone modifications and is bound to regulatory regions (Nalabothula et al., 2015; Ogino et al., 2007). Moreover PARP1 has been shown to associate with p300 and the Mediator complex (Hassa et al., 2001, 2005; Pavri et al., 2005). Therefore PARP1 is a good candidate regarding *Shh* region chromatin unfolding as it has been also linked with chromatin loosening of very large regions (Tulin and Spradling, 2003; Tulin et al., 2002, 2003) with its spreading linked with nucleosome loss at heat shock loci (Petesch and Lis, 2008, 2012).

I therefore examined the effect of fusing PARP1 to my previously designed Tale (Figure Ch5-11A).

Targeting PARP1 to the *Shh* promoter or to its brain-enhancers leads to inconsistent *Shh* activation but does lead to chromatin unfolding only when targeted to the *Shh* regulatory region (on SBE6 or SBE2) but not to the *Shh* promoter (Figure Ch5-11B&C and Table 14). As previously shown with Tale-SIX3, PARP1 mediated chromatin unfolding could be blocked by tNE-CTCF (Figure Ch5-11B & Table 14).

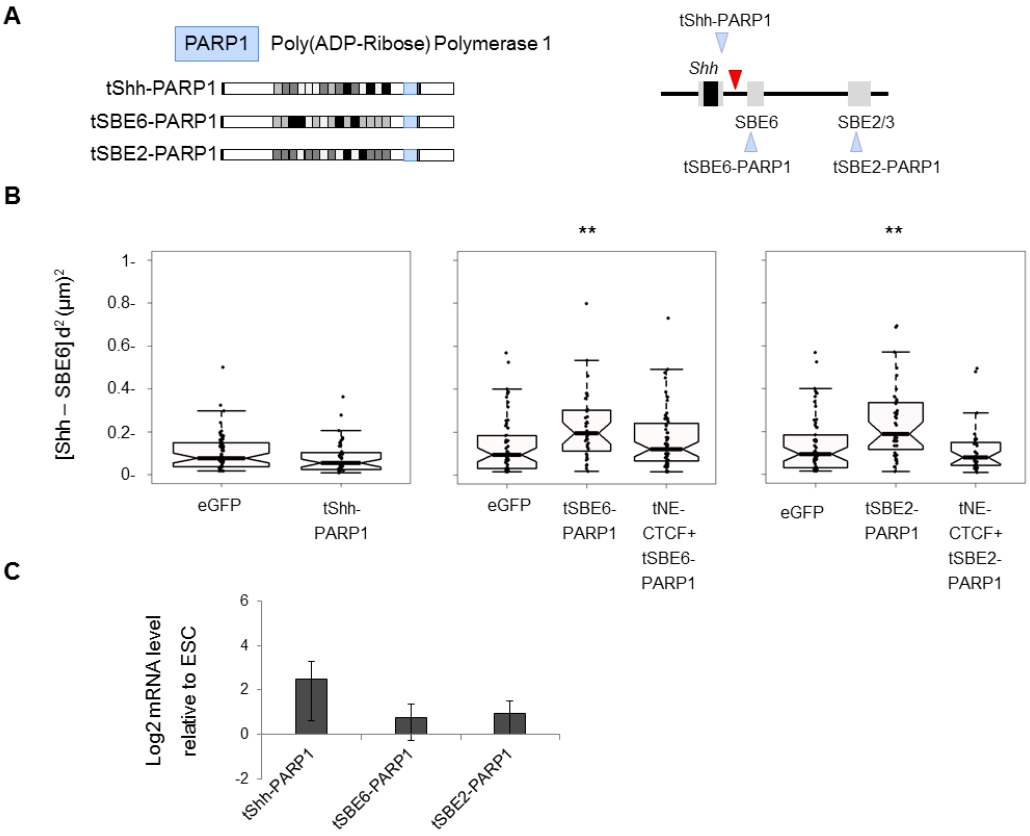

Figure Ch5- 11 Poly(ADP-rybosyl)ation mediated by PARP1 recapitulates the specific chromatin unfolding.

A) Tales-PARP1 constructs and targeting map. B) Boxplots representing Shh-SBE6 squared interprobe distances after Tale-PARP1 transfections. Asterisks on FISH data represent Mann-Whitney U test significance between Tales and eGFP transfected populations, ** for p-values <0.01. C) qRT-PCR showing medians (± s.e.m) of log2 mRNA levels of *Shh* relative to *Gapdh* after Tale-PARP1 transfection in ESC in five biological replicates normalized to eGFP transfection.

	Shh-SBE6	Shh-SBE2/3
tShh-PARP1	2.65E-01	
	6.30E-02	
tSBE6-PARP1	1.23E-03	5.86E-01
tSBE2-PARP1	5.54E-04	6.48E-01
	2.38E-05	
tNE-CTCF+tSBE6-PARP1	1.06E-01	
tNE-CTCF+tSBE2-PARP1	8.83E-01	

Table 14 P-values from eGFP and Tale-PARP1 transfected ESC Mann-Whitney U tests on FISH distances distribution

Exact p-values of Mann-Whitney U tests significance between ESC and Tale-PARP1 revealing the number of replicates.

The role of PARP1 in retinoic acid receptor-mediated gene activation is independent of its catalytic activity (Pavri et al., 2005). To determine whether the catalytic activity of PARP1 is required for chromatin unfolding, I treated Tale transfected cells with Olaparib – a known PARP inhibitor- (Bundred et al., 2013).

Using this drug treatment on tSBE2-SIX3 or tSBE2-PARP1 transfected cells prevented specific chromatin unfolding on treated cells. The treatment only prevented Shh-SBE6 distance increase to occur but did not reduced all chromatin distances in general as Shh-SBE2/3 distances remained the same (Figure Ch5-12A & Table 15).

I could also show that in NPC day seven, a treatment of 1h with Olaparib also prevented the distance increase, as if the chromatin structure established upon differentiation collapsed after Olaparib treatment, without altering Shh-SBE2/3 distances (Figure Ch5-12B & Table 15). Therefore the catalytic activity of PARP1 is essential for maintaining the chromatin unfolding between Shh and SBE6 in NPC.

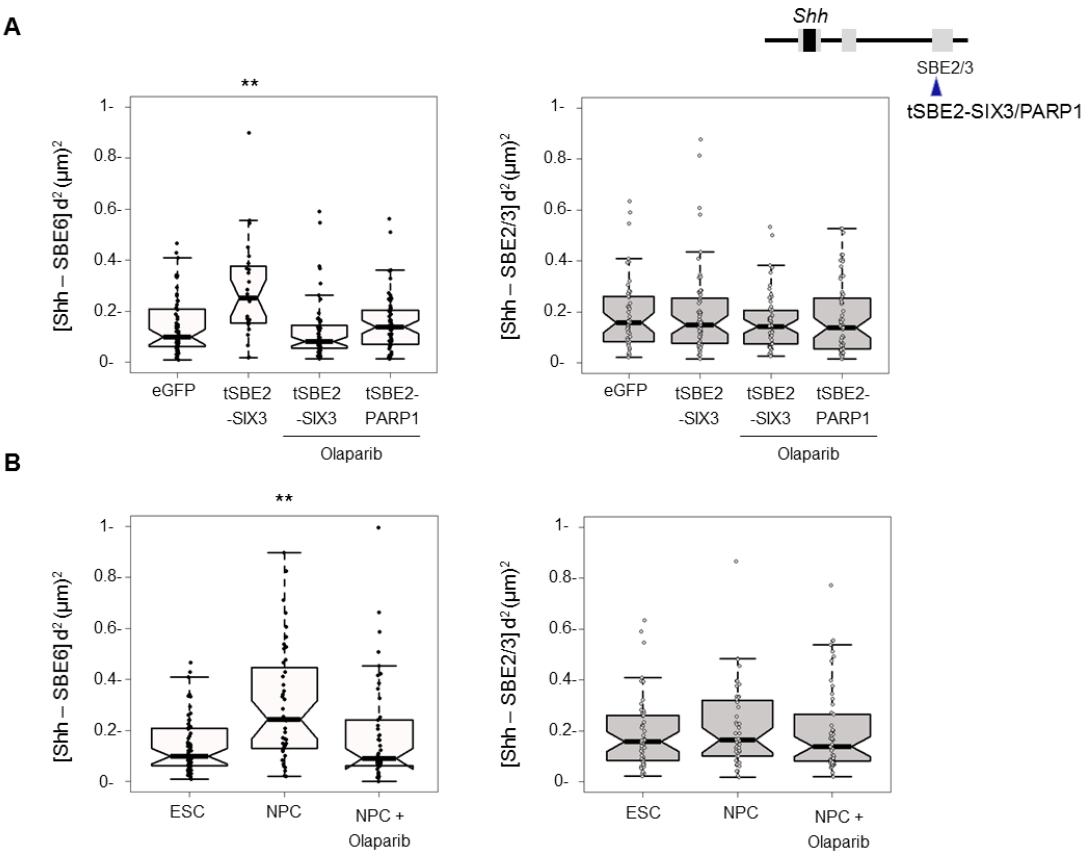

Figure Ch5- 12 SIX3 mediates chromatin unfolding in NPC via PARP1 catalytic activity.

A) Left, Boxplots representing Shh-SBE6 squared interprobe distances for tSBE2-SIX3 with and without Olaparib treatment and tSBE2-PARP1 after Olaparib treatment as a control. Right, Boxplots representing Shh-SBE2/3 squared interprobe distances for tSBE2-SIX3 with and without Olaparib treatment and tSBE2-PARP1 after Olaparib treatment. B) Left, Boxplots representing Shh-SBE6 squared interprobe distances for NPC with and without Olaparib treatment. Right, Boxplots representing Shh-SBE2/3 squared interprobe distances for NPC with and without Olaparib treatment. Asterisks on FISH data represent Mann-Whitney U test significance between Tales or NPC and eGFP transfected or ESC populations, ** for p-values <0.01.

	Shh-SBE6	Shh-SBE2/3
tSBE2-SIX3 Olaparib	2.04E-01	3.13E-01
	6.29E-01	3.96E-01
	9.97E-01	9.67E-01
tSBE2-PARP1+Olaparib	4.17E-01	2.66E-01
	8.16E-01	3.16E-01
	6.80E-01	7.64E-01
NPC Day7 Olaparib	7.99E-01	9.03E-01
	5.46E-01	1.00E+00
	9.46E-01	5.10E-01

Table 15 P-values from eGFP and Olaparib treated Tale-SIX3/Tale-PARP1 transfected ESC or NPC Mann-Whitney U tests on FISH distances distribution

Exact p-values of Mann-Whitney U tests significance between ESC and Tale-PARP1 revealing the number of replicates.

The Belmont peptide is a small acidic peptide DELQPASIDP that can decondense chromatin without triggering gene transcription (Carpenter et al., 2005) or RNA polymerase II recruitment (Therizols et al., 2014). It has been previously successfully fused to Tales to target chromatin decondensation at *Nrp1* and *Ptn* loci in ESC (Therizols et al., 2014). However, the mechanisms of action of the Belmont Peptide remains unknown. To assess if the effect of the Olaparib treatment are specific to PARP1 in the Shh-SBE6 locus, I recruited the 'Belmont Peptide' (BP) (Figure Ch5-13A). Unlike PARP1, the Belmont Peptide led to a Shh-SBE6 nuclear distance increase when targeted to *Shh* promoter or SBE2 confirming that the ability of PARP1 to unfold the chromatin only when targeted to Shh regulatory region is specific (Figure Ch5-13B & Table 16). Moreover Olaparib treatment is specific to SIX3 and PARP1 as Belmont Peptide mediated decondensation is not altered upon treatment (Figure Ch5-13B & Table 16).

Taken together, the Olaparib treatment experiments confirm the role of PARP1 in Shh-SBE6 chromatin unfolding upon SIX3 binding during neural differentiation.

Now that PARP1 seems to regulate this chromatin unfolding using its catalytic activity, it seems that this chromatin decondensation upon gene activation is analogous to the one previously observed in *Drosophila* polytene chromosomes where PARP1 mediates the chromatin loosening of the polytene puffs (Tulin and Spradling, 2003).

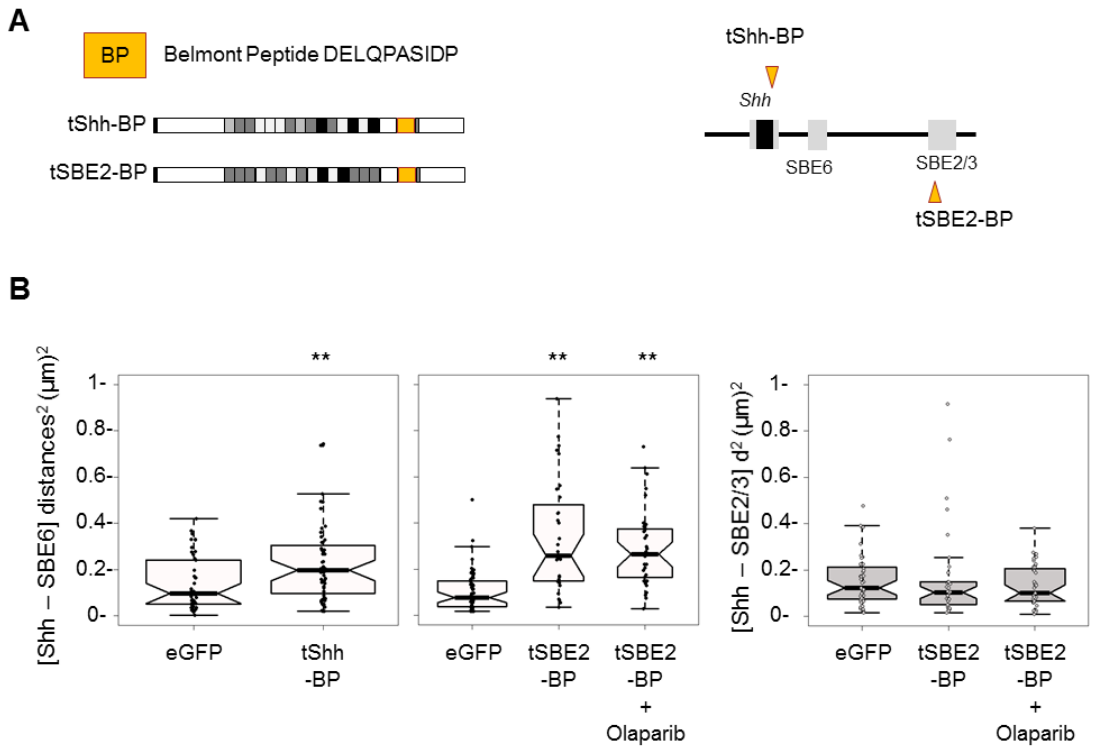

Figure Ch5- 13 Belmont peptide decompaction is not disturbed by Olaparib.

A) Tales-BP constructs and targeting map. B) Left, Boxplots representing Shh-SBE6 squared interprobe distances after tShh-BP or tSBE2-BP transfections with and without Olaparib treatment. Right, Boxplots representing Shh-SBE2/3 squared interprobe distances for tSBE2-BP with and without Olaparib treatment. Asterisks on FISH data represent Mann-Whitney U test significance between Tales- BP and eGFP transfected or ESC populations, ** for p-values <0.01.

	Shh-SBE6	Shh-SBE2/3
tShh-BP	3.05E-03	
tSBE2-BP	1.40E-07	2.90E-01
	2.06E-09	
tSBE2-BP Olaparib	3.12E-05	2.33E-01
	7.46E-10	3.84E-01

Table 16 P-values from eGFP and Tale-BP transfected ESC Mann-Whitney U tests on FISH distances distribution

Exact p-values of Mann-Whitney U tests significance between ESC and Tale-BP treated or not with Olaparib revealing the number of replicates.

5.2.5 State of the intervening chromatin

The Shh-SBE6 chromatin unfolding I have described has been mainly illustrated using 3D-FISH with two fosmid probes, using the inter-probe distances as a measurement of chromatin compaction. Parameters associated with these distances can give some information about the state of the intervening chromatin polymer (Sachs et al., 1995). However, this does not describe thoroughly the state of the intervening chromatin. I was expected from the various ChIP-on-chip for histone modifications that I will learn more about the state of the chromatin. To try to get more detailed information in the nature of the Shh-SBE6 chromatin, I used 2D FISH with Shh probe together with a BAC probe spanning the Shh regulatory region approximately from where NE is located (50kb upstream of Shh) and until SBE4 –thus spanning SBE6 (Figure Ch5-14A). It seems a good region to monitor as it is Shh-SBE6 and Shh-SBE4 distances that increase and that might undergo less interaction according to the various 5C heat-maps.

I observed four types of signal, either a single point (as a normal FISH signal is), a double point, a string (an extended signal) that could be a sign of a rather linear chromatin decondensation and a puff which is represented usually by several dots (Figure Ch5-14B). The quantification of 2D-FISH signals for ESC, tSBE6+tSBE2-Vp64 transfected ESCs (Vp64s that lead to the chromatin unfolding), tSBE2-PARP1 (PARP1) and tSBE2-PARP1 with Olaparib treatment (Olaparib) showed a striking increase of "puff" signals in conditions that undergo the chromatin unfolding i.e. Vp64s and PARP1 (Figure Ch5-14C & Table 17). As the double signal and string signal are constant amongst the different conditions, it is clear that the chromatin unfolding switch from a proper single signal to a puff signal. Therefore it is legitimate to consider this chromatin unfolding similar to the polytene puff observed in *Drosophila*.

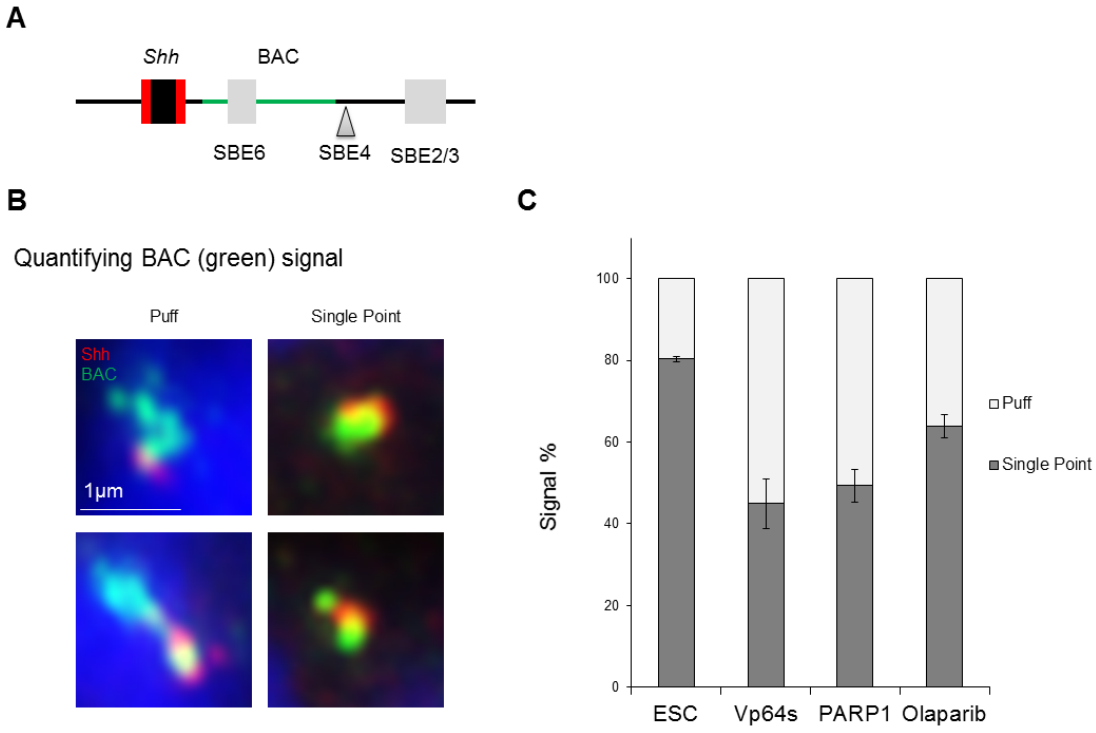

Figure Ch5- 14 2D FISH of the intervening chromatin.

A) Scheme showing the position of the BAC RP24-323C22 spanning chr5: 28,832,721-29,003,817 in mm9 with arrowheads indications SBE6 and SBE4 enhancer positions. B) 2D FISH with Shh fosmid probe in red and the BAC directly labelled in green, the FISH signal can be seen in four different states, either a single point, a double point both categorized as “single point” or the signal is as a string shape or several dots assimilated to a “puff”. C) Quantification of the different states of FISH signals (technical triplicate) between ESC, tSBE6+tSBE2-Vp64 transfected ESCs (Vp64s), tSBE2-PARP1 (PARP1), and tSBE2-PARP1 treated with Olaparib (Olaparib). N=200 alleles.

	ESC	Vp64s	PARP1	Olaparib
Total Counts Single points	527	243	229	356
Total Counts Puffs	130	285	223	212
p-value to ESC	X	0	0	0
p-value to Vp64	X	X	0.158423	0
p-value to PARP1	X	X	X	0.00013

Table 17 Fisher tests significance between ESC, tSBE6+tSBE2-Vp64 (Vp64s), tSBE2-PARP1 (PARP1) and Olaparib treatment on tSBE2-PARP1 transfected cells (Olaparib).

5.3 Conclusions

In this Chapter, I first showed that the Shh-SBE6 chromatin unfolding can be recapitulated by recruiting endogenous factors such as Mediator, p300 and SIX3 to the Shh-Brain-Enhancers.

After observing that synthetically targeting CTCF to the intervening region impaired the Shh-SBE6 chromatin unfolding, I investigated a possible long-range spreading mechanism and which know chromatin remodeller could be responsible for this large restructuration of the chromatin.

I looked at supercoiling, spreading of known chromatin factors like p300 or RNA polymerase II or removal of Polycomb marks and none of those well-known remodeller seem to be the primary determinant of the Shh-SBE6 chromatin unfolding I described.

By looking at other chromatin binding proteins linked with massive decompaction and oligomerization or spreading, I isolated PARP1 as a potential candidate. By recruiting PARP1 directly and also blocking its catalytic activity using Olaparib, I showed that the Shh-SBE6 chromatin unfolding is mediated by SIX3 and PARP1 catalytic activity.

Taken together these results bring into light a new mechanism of promoter activation mediated by its enhancer activity. This result might be very specific to this locus in this peculiar environment, but it shows that probably many more diverse mechanisms apply for various promoter-enhancer interactions.

Chapter 6: Shh regulatory region insulation

6.1 Enhancer activation domain and insulation

Insulating enhancer action has been considered to be an important function of higher order genome organization. Especially TADs and TAD boundaries. As discussed in the Chapter1, 3D chromatin conformation could form -as the nucleolus- as self-organising structures that make reactions more efficient by allowing elevated local concentrations.

It seems that whatever the mechanism, insulator elements or chromosomal rearrangement were able to force and direct the activation signal to a precise promoter element, thus, the idea that the activation signal is constrained in a certain area quickly arose. It seems now clear that more than a simple sequence, enhancers operate in a chromatin domain. Strikingly, it has been shown that an insulated enhancer is instead able to activate transcription from a non-insulated promoter localized in the same domain. Which demonstrates that the transcription factors binding ability on the enhancer remains and that the specificity of enhancer action might rely on the chromatin domain itself (Cai and Levine, 1995; Scott et al., 1999)

Therefore boundary or insulator elements seem to define precise chromatin areas where regulatory elements are active. Work on association rates of transcription factors and other DNA binding proteins revealed that kinetics of protein binding can be ruled by the local packaging of chromatin. Indeed in a facilitated diffusion model, transcription factors engage a first step of 3D diffusion-based exploration that depends on physical and chemical properties of the local nuclear environment such as chromatin exclusion volume and will be concentration-dependent and later operate as a second step a local 1D sliding along the chromatin fibre to efficiently search for targets (Chen et al. 2014, Woringer et al. 2014, Normanno et al. 2015).

The role of CTCF as enhancer insulator is quite unclear, although it has been shown to prevent enhancer action (Bell et al., 1999). CTCF is also extensively linked with TAD formation and therefore guiding the wiring of enhancer-promoter communication (Chapter 1). Therefore I decided to use CTCF during my project to insulate the enhancer action. However, in this case the surprising role of Tale-Delta or Tale-CTCF in preventing the chromatin unfolding is quite hard to fully understand.

In my use of Tale-Delta and Tale-CTCF, both protein ensembles acted as insulators, therefore we cannot conclude on the role of CTCF itself but perhaps more on the Tale binding *per se*. Interestingly in this case Tale binding alone - or with CTCF - prevents tracking or spreading of the chromatin unfolding i.e. the probable parylation of the chromatin but not tracking of activators as previously described. So maybe in this case, the Tale or CTCF action is not representing a classic insulation of the enhancer activity as previously postulated e.g preventing tracking of the activation signal.

I also wanted to test whether the decompaction induced by the Belmont Peptide, which is known to be induced by coactivator proteins such as histone acetylases recruitment via the activators' short acidic-hydrophobic peptide motifs (Carpenter et al., 2005; Tumbar et al., 1999), could be also stopped by CTCF. Using tNE-CTCF and tSBE2-BP, I could see that the decompaction led by the small acidic peptide is not disturbed by CTCF binding (Figure Ch6-1 & Table 18). This data is consistent with the lack of evidence for large histone acetylases recruitment and spreading upon differentiation or Tale-Vp64s targeting in the Shh regulatory region (Chapter 5).

Therefore the insulator role of tNE-CTCF during the chromatin unfolding remain quite surprising.

CTCF insulator activity can be regulated by poly(ADP-ribosyl)ation via PARP1 (Farrar et al., 2010, 2011; Guastafierro et al., 2013). Furthermore CTCF poly(ADP-ribosyl)ation imparts its chromatin insulator properties as lack of PAR due to 3AB treatment abrogated its insulator function (Yu et al., 2004). Using tNE-CTCF and tSBE2-PARP1 or tSBE6-PARP1 was enough to prevent chromatin unfolding perhaps due to CTCF acting as a sink for parylation or acquiring an increased insulation property. Unfortunately the ability of Tale-Delta to also act as insulator in this context remains unexplained.

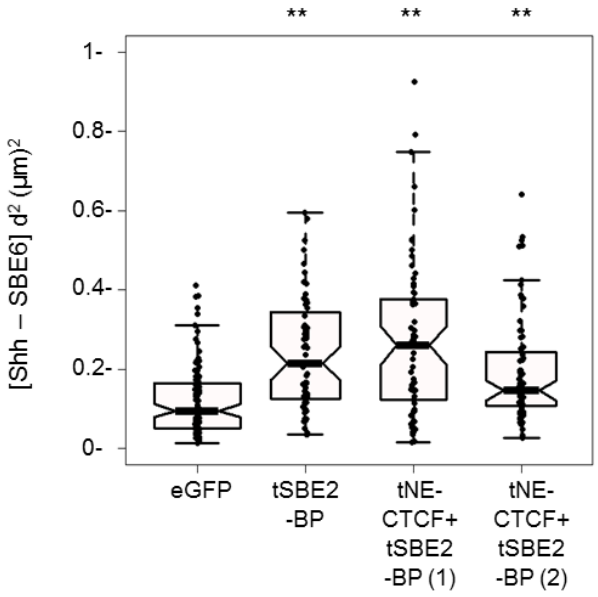

Figure Ch6- 1 CTCF does not impact chromatin decondensation upon Belmont peptide targeting.

Boxplots representing Shh-SBE6 squared interprobe distances after tSBE2-BP with and without tNE-CTCF transfections for two biological replicates. Asterisks on FISH data represent Mann-Whitney U test significance between Tales- BP and eGFP transfected or ESC populations, ** for p-values <0.01.

tNE-CTCF+tSBE2-BP	3.66E-08
	1.33E-05

Table 18 P-values from eGFP and Tale-CTCF and Tale-BP transfected ESC Mann-Whitney U tests on FISH distances distribution

Exact p-values of Mann-Whitney U tests significance between ESC and Tale-BP revealing the number of replicates.

6.2 Shh TAD boundary

Chromatin has been described as segmented into topologically associated domains (TADs) that span several megabases and that are segregated by insulator elements as discussed in Chapter 1. TADs have been described to play an important role in regulating long-range enhancer promoter interplay (Lupiáñez et al., 2015). By disrupting TAD boundaries, inverting CTCF sites sense and creating chromosomal rearrangements, enhancer and promoter communication are completely rewired and this leads to ectopic gene activation that results in pathogenic phenotypes (Gómez-Marín et al., 2015; Guo et al., 2015; Lupiáñez et al., 2015; Tsujimura et al., 2015). This leads to the suggestion that more than specific sequence interactions, a specific chromatin domain is necessary for a proper enhancer-promoter communication (Symmons et al., 2014). This is consistent with the ability of enhancers to act on reporters integrated throughout a TAD (Anderson et al., 2014). Therefore TAD boundaries are thought to be important for constraining enhancer/promoter interactions.

Enhancers are docking site for transcription factors and can activate any responsive i.e. not actively repressed promoter located in a defined chromatin domain. Therefore bringing a strong activator such as Vp64, that recruits the whole co-activator machinery that further leads to the gene expression, into a defined chromatin domain could concomitantly activate surrounding genes that are located in the same domain.

Chromatin conformation studies (5C and HiC), allow us to visualize two overlapping "mini-TADs" located in the main Shh TAD, the first contains *Shh* and goes until just after SBE6, the second one starts at SBE4, contains *Rnf32* and finishes with the same boundary as the main TAD (Figure Ch6-2A) (Dixon et al., 2012, 2015; Rao et al., 2014). Following the TAD (and "mini-TAD") boundary models, it would be expected that the activation effects of Tale-Vp64 should not be spread beyond the TAD boundaries into adjacent TADs. Thus I speculated that as SBE2 is located in the overlapping domain of these "mini-TAD", it could potentially activate *Shh* and *Rmf32* whereas targeting Tale-Vp64 to NE and SBE6 that are strictly confined in the "mini-Tad" containing *Shh* would not activate *Rnf32*.

To assess this, I designed new Tales-Vp64 that are contiguous in the Shh TAD and that go beyond its boundaries to the other neighbouring TADs (Figure Ch6-2A). The Shh TAD

possess only two genes - *Shh* and *Rnf32* - located at the opposite ends. CRTL *En2* and *Rmb33* are located on the next TAD beyond *Shh*, whereas *Nom1* is in the neighbouring TAD on the opposite side, close to *Rnf32*. I first looked at the profile of gene activation induced by Tales-Vp64 targeted to these genes in these three TADs in ESC (Figure Ch6-2B). It is important to note that only *Shh* and *En2* are off in ESC. Thus the very low levels of activation created by tRbm33-Vp64, tRnf32-Vp64 and tNom1-Vp64 to their own target could be explained by the fact that these genes are already active and it is then harder to activate them even more.

The qRT-PCR data highlighted two interesting results. It first appeared that only Tale-Vp64s targeted between *Shh* and SBE2 could strongly activate *Shh*. Secondly, as predicted, only tSBE2-Vp64 was able to lead to *Rnf32* activation to levels as high as those achieved by Tale-Vp64 targeted to the *Rnf32* promoter.

This data also shows that *Shh* and the non-coding RNA *9530036O11Rik* are the easiest transcripts to induce by a transfection of a Tale-Vp64 to the Shh regulatory region. Indicating perhaps that the *Shh* chromatin environment could allow *Shh* to be expressed easily when strong activators come into its vicinity. This would be consistent with the need of a highly compacted conformation of the Shh locus and a chromatin unfolding model that enhances *Shh* transcription upon activator presence.

Comparing qRT-PCR and 5C heat-maps allow us to draw a very clear parallel between these data (Figure Ch6-2A). Thus this seems to confirm that one Tale-Vp64 off-target activation might depend on the close chromatin conformation the target lies in. This would also be consistent with the view of an enhancer being a docking site for an activator and thus rewiring enhancer activity when moving the enhancer outside a “mini-TAD” which is persistent with the view of an “enhancer domain” or chromatin hub discussed in Chapter 1.

It is also interesting to note that tRnf32-Vp64 and tNom1-Vp64 seem to bypass their boundary more easily. This would need further investigation on the nature of the TAD boundaries in order to draw a conclusion.

To further dissect this, it would be interesting the delete the TAD boundary or remove the CTCF site and change their orientation to see if the domain of activation that Vp64 can reach differs.

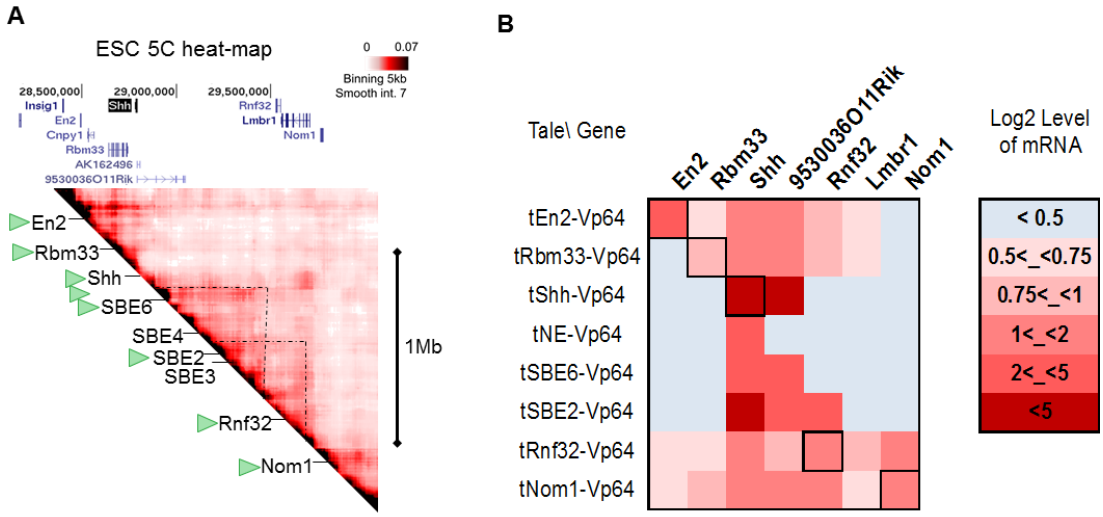

Figure Ch6- 2 Tale-Vp64 activates the surrounding chromatin.

A) ESC 5C heatmap (chr5:28317087-30005000) with 5kb binning and smoothing with the median of 7 surrounding interactions frequencies displaying Tale-Vp64 targeting represented as green triangles. B) Log2 mRNA levels for *En2, Rbm33, Shh,* long non-coding RNA 9530036O11Rik, *Rnf32, Lmbr1 and Nom1* qRT-PCR after Tales-Vp64 are displayed as a log2 heatmap.

Chapter 7: Discussion

7.1 Distal Shh-Brain-Enhancers regulate chromatin unfolding to promote *Shh* expression upon neural activation

Studying *Shh* and its brain-enhancers communication during neural differentiation revealed a previously undescribed mechanism of enhancer-promoter communication. I was first able to establish two new enhancers that are active in NPC, with only one of them (which I named SBE6) a robust *Shh* activator in the forebrain (Chapter 2). I was then able to observe that upon neural differentiation the *Shh* promoter moves further away from SBE6 but not SBE2. A nuclear distance increase that is very specific to the Shh-SBE6 region and arises from day 3 onward is incompatible with the common long-range chromatin-looping model. This distance increase is general and homogenous through the population of cells that are homogeneously expressing *Shh*. Deleting SBE6 was enough to reduce *Shh* expression in NPC and abolish the distance increase, revealing that this neural chromatin unfolding is linked with enhancer activity. Furthermore this effect is specific to SBE6.1, as deleting SBE6.2 did not influence expression or chromatin structure in NPC.

Subsequently I was able to show that the imaging system I used – super-resolution SIM and FISH – allows me to visualize a chromatin loop if a looping event occur. And using a synthetic biology approach, I could demonstrate that the chromatin-unfolding event does not simply result from activating *Shh* directly but that it results from a co-activation of SBE6 and SBE2. The chromatin unfolding occurs after a certain concentration of activator protein is loaded onto the Shh regulatory region. The opening of the *Shh* region is associated with a general increase in *Shh* expression across cells in the population. Preventing this opening by deleting SBE6 or synthetically blocking it by targeting CTCF to the intervening region impaired *Shh* expression. Therefore the underlying mechanism is linked with a long-range spreading-like.

I later established that the recruitment of SIX3 onto SBE2 is critical for the chromatin unfolding to occur and that SIX3 mediated chromatin unfolding is able to enhance *Shh* activation via its Brain-enhancers, which is consistent with the results from NPC day 3 and

SBE6 -/- NPC day 7 that are expressing *Shh* at low levels but are not displaying the chromatin unfolding.

Taken together, it seems that preventing the chromatin unfolding negatively affected *Shh* expression levels and that promoting chromatin unfolding with SIX3 positively affected *Shh* activation when co-activators are present. These results illustrate the strong interplay between chromatin unfolding and robust *Shh* activation. This leads us to postulate that in holoprosencephaly cases (HPE) caused by impaired SIX3 binding, *Shh* haploinsufficiency could result from abrogated chromatin unfolding (Belloni et al., 1996; Geng et al., 2008; Roessler et al., 1996; Wallis et al., 1999).

I next demonstrated that SIX3 chromatin unfolding is mediated by PARP1 poly(ADP)ribose activity to open up the Shh-SBE6 region, as Olaparib which blocks PARP1 catalytic activity was able to prevent the specific Shh-SBE6 chromatin unfolding in NPC and ESC transfected cells.

These results help me to draw a model describing the chromatin architecture of the Shh regulatory region upon neural differentiation from ESC to NPC and *Shh* enhancer driven activation (Figure Ch7-1). In ESCs the Shh regulatory region is in a closed conformation and during early stages of differentiation activators lead to induction of *Shh* without the prior opening of the Shh-SBE6 region. In that first condition, the cell express *Shh* at low levels. This is observed in early differentiation to NPCs, in SBE6 deleted NPC, in tNE-CTCF transfected ESC and I speculate in SIX3 disrupted cases of HPE. With the full complement of activators loaded onto the Shh regulatory region and SIX3 recruitment followed by poly(ADP-ribosyl)ation of the region by PARP1, the chromatin unfolds and enhances *Shh* expression. This chromatin unfolding may make the region more accessible to factors that facilitate gene induction.

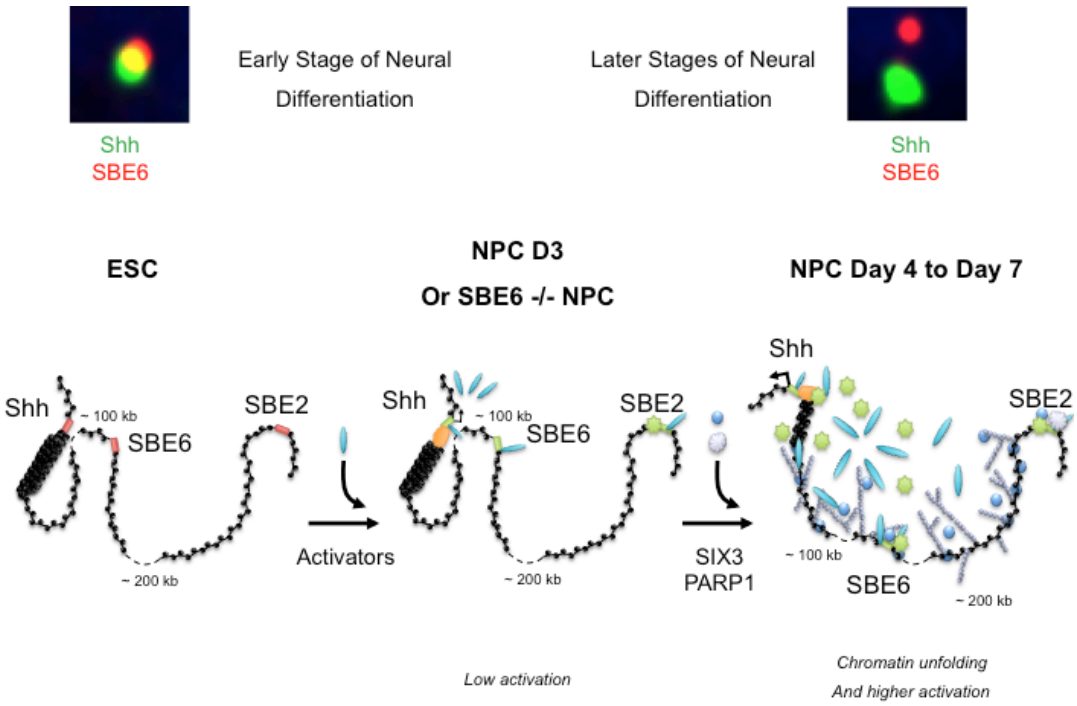

Figure Ch7- 1 Shh-Brain-Enhancers promote a chromatin unfolding in Shh regulatory region upon a neural activation.

In ESC, the Shh regulatory region is in a closed conformation as illustrated with the Shh-SBE6 FISH signal, and upon early stages of differentiation activators lead to Shh activation without yet opening the Shh-SBE6 region as observed in early NPC differentiation, in SBE6 deleted NPC or in tNE-CTCF transfected ESC. All those conditions display a low Shh activation. With SIX3 recruitment and by poly(ADP-ribosyl)ing the region with PARP1, the chromatin finally unfolds which then enhances *Shh* expression.

7.2 New insight on enhancer mechanisms of action?

7.2.1 An assembly of transcription factor motifs

Functional characterization of enhancers is often realised through the property of enhancers to be able to drive expression of a minimal reporter gene in an episomal vector - thought to lack physiological chromatin - or inserted as a transgene into a host genome such as zebrafish or mouse. By comparing chromosomal and episomal activity of enhancers, it seemed that the chromatin environment led to a much cleaner representation of the enhancer's activity (Inoue et al., 2016). However, we cannot deny that a reporter assay performed on a non-integrated episomal vector gives a good and rapid indication of the enhancer pattern activity. This reveals that primarily an enhancer can be summarized as a cluster of tissue-specific transcription factor binding sites that are the key definers of cell type and temporal enhancer activity. To a certain extent, this view can be illustrated by a recent study based on motif analysis and not sequence conservation to find new enhancers with similar activities. This analysis revealed that sequences required for enhancer function work in assemblies of transcription factor motifs, and that the order does not matter as only the presence of all those motifs is required the pattern activity of the enhancer (Yao et al., 2016).

This idea is similar to the initially popular view of enhancers as an extension of the promoter. Transmission of transcription machineries or signals to the neighbouring minimal reporter promoter in an episomal vector is the critical characteristic that all regulatory element share. As a docking site for tissue-specific transcription factors, enhancers serve as bidirectional entry sites for either RNA polymerase II or one of its subunits. This first provides an explanation regarding their position and orientation independence. It also implements the first definition of the tracking model. After serving as an anchoring site, tissue-specific transcription factors track along the DNA sequence or the chromatin fibre with insulator elements setting boundaries of the activation domain.

7.2.2 Three main mechanisms?

Regarding nearby enhancers (few kb), it is commonly accepted that their proximity to their target promoter is enough to induce activation after transcription factor binding and transcription machinery recruitment. It is generally agreed that perhaps a short tracking might occur to reposition the machinery on the transcription start sites, but only limited mechanistic dissection has been done on these enhancers in mammals.

For distal enhancers the looping model prevails, in which the spatial co-localisation of the promoter and enhancer seems crucial in order to concentrate transcription factors at the anchoring site. It has been recently established that direct promoter-enhancer interaction is responsible for driving allele specific transcriptional bursting (Bartman et al., 2016; Fukaya et al., 2016).

Regarding *Shh*, it is interesting to consider that the ZRS located 800kb upstream of Shh is in a permanent close conformation with Shh promoter and becomes even closer upon activation (Williamson et al., 2016). This subtle 3D repositioning is already different from other loci that undergo drastic "on/off" state of looping such as the β-globin LCR where the enhancer is positioned far from the promoter when inactive (Simonis et al., 2006; Tolhuis et al., 2002). On the contrary, *Shh*-ZRS, similar to the HoxD locus, are permanently located very close to each other, whether the gene is expressed or not and undergo co-localisation upon activation (de Laat and Duboule, 2013; Montavon et al., 2011). The stable loop conformation seems to correlate with paused RNA Polymerase II occupation and release upon direct looping and transcription initiation (Ghavi-Helm et al., 2014). Could we consider those two types of looping mechanisms similar? It seems that in one case the physical forces needed for accomplishing the enhancer-promoter chromatin loop will be greater as the initial distance between enhancer and promoter is bigger. We can also speculate that is one case in which the chromatin conformation should allow a dynamic movement whereas the other case would be in a tighter controlled conformation involving other types of protein binding to maintain the locked position.

The chromatin unfolding model I describe for Shh-Brain-Enhancer seems to enhance activation through the recruitment of activators and it could be seen as another type of facilitated tracking model for a locus that is highly compacted. It is fascinating to realize that

in the same locus Shh has different enhancers that utilise different mechanisms of regulation. The main question that remains is what determines the mechanism that the enhancer will follow. Is it a co-evolution of the transcription factors and the chromatin environment? To answer that it would be interesting to use genome engineering to swap the positions of ZRS and SBE6 or SBE2 and see if ZRS/Shh is still following the looping model or if the chromatin unfolding will take the lead. Unfortunately this experiment will be difficult to perform as no *in vitro* cell line model exist to study the ZRS, all the studies have to be performed in mouse limb tissues, and reciprocally it is quite hard to pin down what tissues and stage NPCs represent in the developing mouse brain. However, we could test if SBE6.1 inserted into Lmbr1 in 46c cells follows a looping or chromatin unfolding mechanism upon neural differentiation.

Moreover for brain activation of *Shh* in NPCs it seems that more than one enhancer is involved, whereas for the precise limb expression of Shh only the ZRS has been described so far. Therefore we should also consider "the strength" of an enhancer, if the activation is due to a composition of various enhancer activity or due to a single specific strong enhancer (Bothma et al., 2015). And if that interferes with the mechanism used for enhancer action? Indeed we could imagine that if the activation relies on a unique distal enhancer, it would be better if it is already located near the promoter in a closed conformation with paused polymerase awaiting to be released. However if it works with a cooperation of several enhancers, depending on the chromatin context it will used a tracking model, or a dynamic looping or even a facilitated tracking model where as a step-by-step process, proteins scan the chromatin as a preliminary step until the cognate promoter is met and that a stable loop is maintained when the tracking is complete. The facilitated tracking model observed the unidirectional spreading of acetylated histones H3 and H4, histone acetyltransferases CBP/p300, P/CAF, subunits of the human SWI/SNF or RNA polymerase II and TBP that synthesized short, polyadenylated, intergenic RNAs (Gribnau et al., 2000; Hatzis and Talianidis, 2002; Kim and Dean, 2004; Masternak et al., 2003; Spicuglia et al., 2002; Zhao and Dean, 2004; Zhu et al., 2007).

Only the case-by-case study of precise enhancer-promoter interaction will help us to have a clearer view on how many different mechanisms exist and what determines the model, the chromatin conformation, the number of enhancers involved and their position, the TAD position or the gene position within the TAD?

But it might take more time as many enhancers are still awaiting annotation. Recent work from the Stark lab estimated between at least 50,000 to 100,000 developmental enhancers in the 170-megabase *Drosophila melanogaster* genome which suggests that the 3-gigabase human genome could contain up to several million enhancers (Kvon et al., 2014) as even only Shh seems to contain at least 30 more unannotated brain enhancers (Chapter 2).

7.3 Characteristics of the chromatin unfolding

The new mechanism I described during my thesis is so far very specific to the Shh locus in the context of the neural differentiation. However, we do not know whether this mechanism can be generalizable for other loci and what could be the rule for following that pathway of enhancer-promoter communication, as opposed to chromatin looping.

I first tried to decipher which characteristics made the Shh locus special. *Shh* is expressed in neural tissues and possess many brain enhancers and with many more yet to be annotated (Chapter2). Also, Shh is located at a TAD border and near a large gene desert where its enhancers are positioned. Furthermore Shh is a strong Polycomb target. The Polycomb marks do not disappear upon gene expression and in ESC Ring1b$^{-/-}$ (Polycomb complex mutant), *Shh* is not expressed whereas many other Polycomb target become expressed following the removal of the repressive mark (Rob Illingworth, *unpublished*). This could potentially indicate that Shh possess several levels of repression to prevent ectopic activation of the powerful morphogen.

I picked other candidates that followed those characteristics (Figure Ch7-2A) and looked at the corresponding HiC data available for mouse ESC and Cortex (http://promoter.bx.psu.edu/hi-c/view.php) from the Ren Lab (Dixon et al., 2012, 2015; Rao et al., 2014). Interestingly, all three loci, Six3, Sp8 and DLX6, seem to lose some degree of interaction in the gene desert where there putative brain enhancers are located in the mouse cortex (Figure Ch7-2B, Ch7-3).

Of course, this large scale decompaction could be due to many other factors and should be further dissected by FISH. However, the fact that the chromatin unfolding at the Shh locus seem to be present in this HiC data in mouse cortex suggests the decompaction of those other loci could be meaningful.

A

Name	code	expression	Brain enhancers	Gene Desert	Tad Border, decompaction in Cortex	polycomb target	Sur expression in Ring1b -/-
Sonic Hedgehog	Shh	neuronal development	yes	yes	yes, yes	yes	no
sine oculis-related homeobox 3	**Six3**	**neuronal development**	**yes**	**yes**	**yes,yes**	**yes**	**no**
trans-acting transcription factor 8	**Sp8**	**Limb & neural dvt**	**yes**	**yes**	**yes, yes**	**yes**	**no**
Homeobox protein DLX-6	**DLX6**	**neuronal development**	**yes**	**small**	**yes, yes**	**yes**	**no**

B

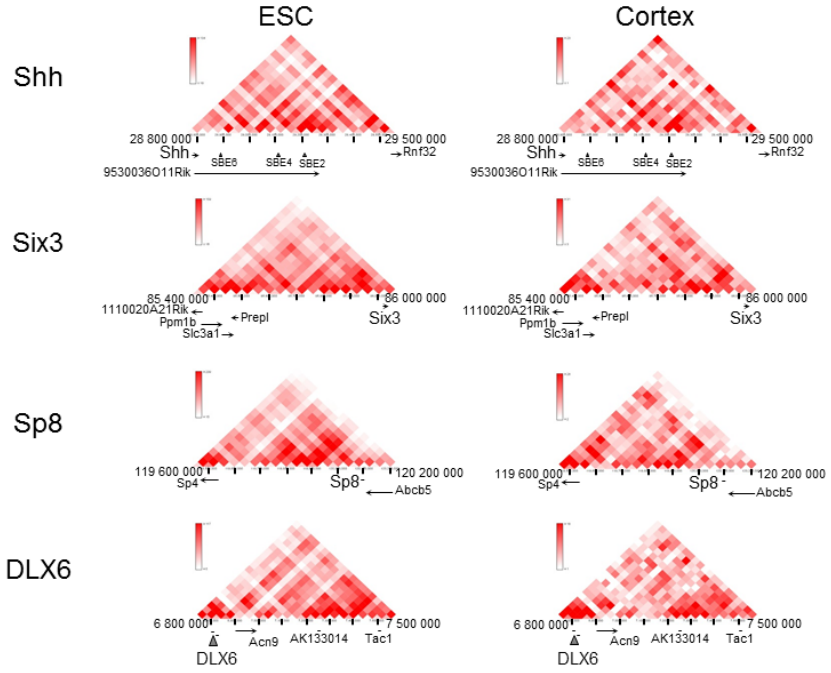

Figure Ch7- 2 Possible loss of interaction at other developmental genes.

A) Table with the characteristic used to pick similar loci. B) HiC data from the Ren Lab in mouse ESC and Cortex for Shh, Six3, Sp8 and DLX6 loci.

Shh

Six3

Sp8

DLX6

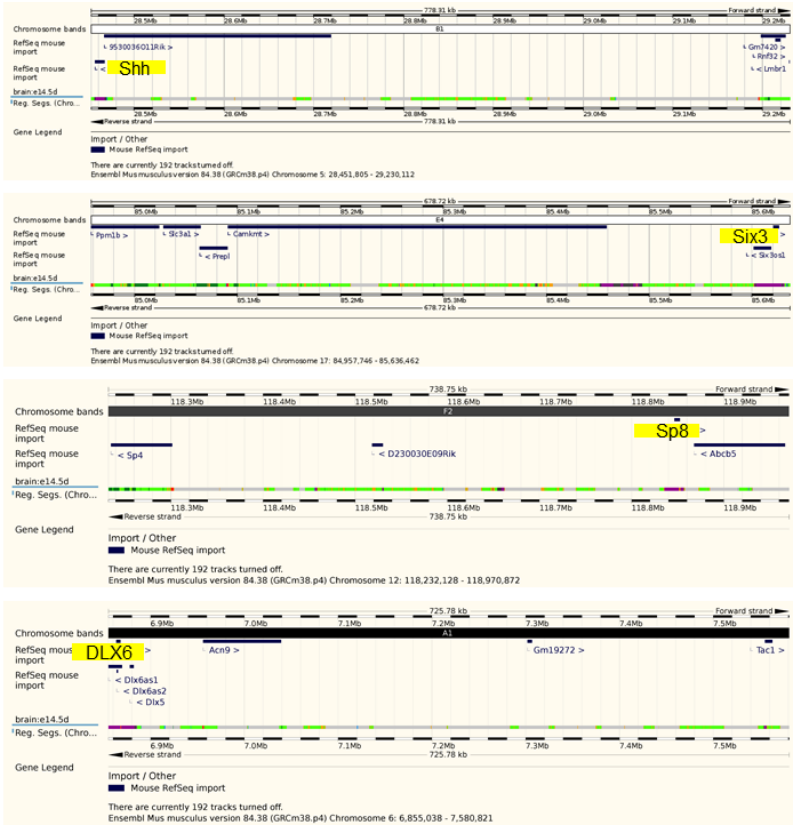

Figure Ch7- 3 Putative Brain enhancers (dark yellow) for Shh similar loci.

7.4 Live-cell imaging revolution

Using fixed samples only gives a static snapshot of chromosome organisation. Having the ability to study the dynamics of chromosome movement of the interactions that form and dissolve during gene regulation in living cells will be a tremendous step forward. Toward that goal, tools to visualize single loci by live cell imaging are still awaiting full development.

In the new mechanism I have been describing where enhancers promote a chromatin puff probably similar to polytene puffs observed in *Drosophila,* the dynamic component of that mechanism is still missing. Although we could see that it appears progressively during neural differentiation and seems perhaps stable as it could be seen also by 5C and generally as a distance shift in the whole population by FISH, a live-cell imaging approach would characterise more the kinetic of this chromatin unfolding.

Using an GFP-tagged dCas9 with an array of 24 structurally optimized sgRNAs tiling along *Shh,* a single endogenous locus should be tracked in a living nucleus (Chen et al., 2013). Similarly, instead of using 24 sgRNAs, a recent study used a protein scaffold that can recruit several copies (24) of an antibody protein fused to a fluorescent dye (superfolder GFP, sfGFP (Pédelacq et al., 2006)). This SunTag complex fused to a dCas9 can be targeted using a single sgRNA and enable live cell tracking of a single locus (Tanenbaum et al., 2014).

Therefore, in theory, combining those two recent imaging technologies using dCas9/CRISPR and Tale system should enable dual colour live visualisation of two isolated loci (Figure Ch7-4). I briefly undertook the cloning of a dCas9-mCherry with 24 sgRNA tilling Shh promoter and upstream gene body. In parallel, I cloned the SunTag protein scaffold to various Tale I built such as tShh, tSBE6 and tSBE2. A transfection of a Tale-SunTag scaffold alongside the antibody fused to a sfGFP (sfGFP-AB) would allow in theory the visualization of the Tale targeted locus.

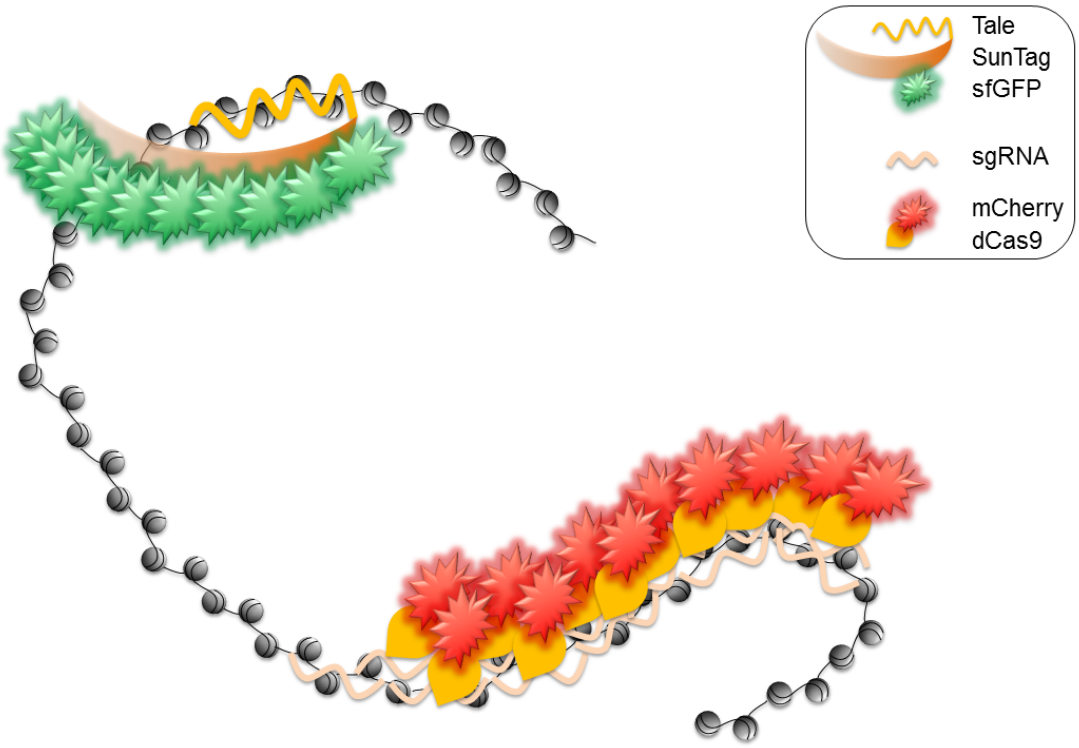

Figure Ch7- 4 Live-Cell Imaging theory.

Two sequences tagged with either Tale fused with SunTag protein scaffold that recruits sfGFP or sgRNA tilling the sequence and recruiting dCas9-mCherry.

After a brief control using the 24 Shh- sgRNA with dCas9-mCherry and tShh-SunTag and sfGFP-AB to monitor proper colocalization, it seemed that the system needs substantial optimisation and will be pursued by other members of the lab.

This system will be likely to be cleaner in a stable line, as the system requires a co-transfection of 27 plasmids (24 Shh sgRNAS, dCas-mCherry, tShh-SunTag, sfGFP-AB) resulting in many cells lacking the full cocktail and therefore the poor number of cells possessing 2 loci to visualise. Furthermore, integrated plasmids will decrease the background and noise as the expression will be lower and cleaner. Unfortunately creating those cell lines and following the enhancer-promoter communication upon differentiation would be another project in its own right, but will be an important development in the future of the enhancer field.

Chapter 8: Materials & Methods

8.1 Cell culture

46c mouse embryonic stem cells (mESCs), derived from E14tg2A, contain a GFP insertion into the Sox1 locus (Ying et al., 2003). 46c and Ring1B$^{-/-}$ ESC were grown in GMEM supplemented with 10% fetal bovine serum (FBS), 1,000 units/ml LIF, nonessential amino acids, sodium pyruvate, 2-β-mercaptoethanol, L-glutamine, and Penicillin/Streptomycin. ESCs were cultured and differentiated into NPCs with N2B27 medium as described previously (Pollard et al., 2006; Taylor et al., 2013). ESCs were transfected and sorted following (Therizols et al., 2014) protocol. Olaparib treatment was performed as follows, 10µM Olaparib (stock 100 mM in DMSO) diluted in ESC or NPC media was added on 1x10^6 cells 24h after transfection or 5h after FACS for NPC, cells were treated for 1h30 and media was changed and cells were fixed for FISH

ESCs were transfected with plasmids using Lipofectamine® 2000 Reagent (Invitrogen cat. N°11668) following the manufacturer recommendations. 1x10^6 ESCs were transfected in a 6-well plate with 2.5µg of circular plasmid and 7µl of Lipofectamine. Medium was changed 5h after transfection. Transfected cells were sorted based on eGFP expression by FACS 24h after transfection and re-seeded on slides or 6-well-plates. To sort GFP^{+ve} cells after transfection or differentiation, cells were trypsinized and resuspended in PBS + 10% medium. Flow cytometric analysis was performed using the 488nm laser of a BD FACSAriaII SORP (Becton Dickinson) with 525/50 nm bandpass filters. BD FACSDiva software (Becton Dickinson, Version 6.1.2) was used for instrument control and data analysis.

8.2 Quantitative analysis of gene expression

RNA was prepared from approximately 1x10^6 46c mESCs using the RNeasy mini kit (Qiagen) according to the manufacturer's protocol, including a DNaseI (Qiagen) treatment for 15 minutes at room temperature. cDNA was synthesized from 2 µg purified RNA with Superscript II reverse transcriptase (Invitrogen) primed with random hexamers (Promega). Real-time PCR was carried out on the Roche LightCycler 480 Real-Time PCR System using a Lightcycler 480 Sybr Green detection kit (Roche) as described previously (Therizols et al., 2014). Primers sets for qRT-PCR were:

Expression RT-qPCR primers	
Gapdh Fw	ATCACCATCTTCCAGGAGCGAG
Gapdh Rv	GACCCTTTTGGCTCCACCCTTC
Oct4 (Pou5f1) Fw	CGAGAACAATGAGAACCTTC
Oct4 (Pou5f1) Rv	CCTTCTCTAGCCCAAGCTGAT
Nestin Fw	GATCGCTCAGATCCTGGAAG
Nestin Rv	AGGTGTCTGCAAGCGAGAGT
Shh Fw	ACGATTTAAGGAACTCACCC
Shh Rv	TTGTCTTTGCACCTCTGAG
En2 Fw	GGTCTACTGCACGCGCTATTCT
En2 Rv	AAACTCAGCCTTGAGCCTCTGG
Rbm33 (Pr8) Fw	GCTTCCTACACAGCCGTCTGTT
Rbm33 (Pr8) Rv	GGAACAGACTTCTGATTGGCTGC
9530036O11Rik Fw	GCCTGAAACACACAGAATGC
9530036O11Rik RV	GATGGGAGAACTCAGCCAAG
Rnf32 Fw	GCAGCTAGATGAAAGGTGTGGC
Rnf32 Rv	AGTGGAGTCAGGCAGATGGAAC
Lmbr1 Fw	GGAGGAAAAAGGCTTCAGCTTGG
Lmbr1 Rv	GTCTCATCAACCAGCAGGCAAAG
Nom1 Fw	GGCTCTGAAGAACAATGACCTGC
Nom1 Rv	AGATGCGAAGCCGAGTCTCAGA
ChIP qPCR primers	
tNE Fw	CAAGAAAACATGGGTCTAGGG
tNE Rv	TGAATTGGAACATTTCTTCAGG
Actin Fw	CCTCGATGCTGACCCTCATCC
Actin Rv	GACACTGCCCCATTCAATGTCTC

Table 19 Primers sets for qRT-PCR

The real-time thermal cycler was programmed as follows: 15 min Hotstart; 44 PCR cycles (95°C for 15 s, 55°C for 30 s, 72°C for 30 s). The relative mRNA expression for each primer set in each sample was measured by the Lightcycler software and normalized to the mean for *Gapdh* from at least 2 biological replicates and technical triplicates. Finally, the log2 of the ratio relative to eGFP transfected ESCs was calculated when mentioned.

8.3 Computational analysis of the SBE6 region

Evolutionary conservation of the SBE6 region was assessed using the "Vertebrate Multiz Alignment & Conservation/Multiz Alignments and Conserved elements" tracks in the UCSC genome browser (Kent et al., 2002). This delineated the following sub-regions for further analysis:
SBE6.1: Chr5:28889688-28890461, SBE6.2: Chr5: 28893935-28895000 (mm9)

rVISTA (Loots et al., 2002) was used to align the mouse and human orthologous sequences, with the default sequence aligner (LAGAN) and default parameters. Transcription factor binding sites (TFBS) for known forebrain TFs (Nord et al., 2015) available on the rVISTA server were selected (Arx, Maf, Dlx5, Pbx1, ER81, Six3, Vax1).

JASPAR (Mathelier et al., 2016) was used independently on the mouse and human core sequences, searching for potential neural activity present in the Jaspar Core Vertebrata matrices list (DLX6, PBX1, ETV1, Six3, SP8, and VAX1) with the default parameters (Relative profile score threshold 80%). Hits were then highlighted on the rVista alignment.

8.4 Zebrafish enhancer reporter assay

The putative SBE6.1 and SBE6.2 enhancer were cloned by PCR amplification of the relevant fragment and flanking sequence from mouse genomic DNA, using Phusion high fidelity polymerase (NEB) and the following primers:

Sbe6.1 Fw B4 :
AGGGGAGAACTTTGTATAGAAAAGTTGGCGCGCCCACCTGCTTCTCTGAGGAA
Sbe6.1 Rv B1R : **AGGGGACTGCTTTTTTGTACAAACTTG**CTTAGGCCATTGTGCCCAC

Sbe6.2 Fw B4 :
AGGGGAGAACTTTGTATAGAAAAGTTGGCGCGCTGAAGTCAAGGGCCTGGTACT

Sbe6.2 Rv B1R : **AGGGGACTGCTTTTTTGTACAAACTTG**ATCAGCCCTCCAGTTTGACT

-veCTRLs

Sbe6.1 Fw B4 :

AGGGGAGAACTTTGTATAGAAAAGTTGGCGCGCCGAGTGCAGGTGTTTGTGAA

Sbe6.1 Rv B1R : **AGGGGACTGCTTTTTTGTACAAACTTG**CCTCAACACAGCATTGCCAA

Sbe6.2 Fw B4 :

AGGGGAGAACTTTGTATAGAAAAGTTGGCGCGCAGAGAGTGAAGATTCCCAGCT

Sbe6.2 Rv B1R :

AGGGGACTGCTTTTTTGTACAAACTTGTGAGGCAGTGTCTATCTTTTGAC

attB4 and attB1r sequences (bold) were included in the PCR primers for use with the Gateway recombination cloning system (Invitrogen, 12538120). The amplified fragment was first cloned into the Gateway pP4P1r entry vector and sequenced using M13 forward and reverse primers for verification. The elements in the pP4P1r vector were combined with a pDONR221 construct containing either a Gata2 promoter-eGFP- polyA or a Gata2 promoter mCherry- polyA cassette (Bhatia et al., 2015), and recombined into a destination vector with a Gateway R4-R2 cassette flanked by Tol2 recombination sites.

Reporter plasmids were isolated using Qiagen miniprep columns and were further purified using a Qiagen PCR purification column (Qiagen), and diluted to 50 ng/µl with DNAse/RNAse free water. Tol2 transposase RNA was synthesized from a NotI-linearized pCS2-TP plasmid using the SP6 mMessage mMachine kit (Ambion), and similarly diluted to 50 ng/µl. Equal volumes of the reporter construct(s) and the transposase RNA were mixed immediately prior to injections. 1–2 nl of the solution was micro-injected per zebrafish (Danio rerio) embryo at the 1- to 2-cell stage for up to 200 embryos. Embryos were screened for fluorescence at 1–5 days post-fertilization i.e. 24–120 hpf (hours post fertilization) and raised to adulthood by Shipra Bhatia. Germline transmission was identified by mating of sexually mature adults to wild-type fish and examining their progeny for fluorescence. F1 embryos from 3–5 F0 lines showing the best representative expression pattern for each construct were selected for confocal imaging. A few positive embryos were also raised to adulthood and F1 lines were maintained by outcrossing. A summary of the number of independent lines analysed for each construct and their expression sites is included in Table

2. Imaging of zebrafish reporter transgenic embryos was carried out as previously described (Bhatia et al., 2015)

8.5 Mouse Enhancer LacZ Reporter assay

For mouse transgenics, the same PCR amplicons were used for the generation of enhancer-reporter constructs and the generation, the blastocysts injection and the analysis of transgenic lines were carried out as previously described (Ravi et al., 2013) injected by the CBS-IGMM Operations and Transgenic Facility. The two E11.5 SBE6.1-LacZ embryos were dissected in PBS and left in LacZ fix for 1h (1% Formaldehyde, 0.2% glutaraldehyde, 2mM $MgCl_2$, 5mM EGTA, 0.02% NP40) then washed twice for 5 min with PBS+0.02%NP40 and embryos were incubated overnight in the dark at 37°C in the X-gal Staining Solution (5mM $K_3Fe(CH)_6$, 5mM $K_4Fe(CH)_6$3H2O, 2mM MgCl2, 0.01% Nadeoxycholate, 0.02% NP40, 0.5 mg/ml X-gal). Embryos were then fixed with 4% PFA in PBS.

8.6 Mouse *In situ* hybridisation

Mouse *in situ* hybridisation was performed with DIG-labelled gene-specific antisense probes as previously described (Hecksher-Sørensen et al., 1998). The *Shh* probe was provided by Andy McMahon (Echelard et al., 1993).

8.7 Deletion of SBE6 from the 46c ESC genome.

46c ESCs deletions were produced using the Crispr/cas9 system. SBE6.1 and SBE6.2 specific gRNA primers were cloned into the cas9 plasmid pX458 following protocols from the Zhang lab (Cong et al., 2013; Ran et al., 2013a, 2013b). 46C ESCs were transfected with the resulting plasmids (2.5µg) using Lipofectamine® 2000 Reagent (7µL) (Invitrogen cat. N°11668) following the manufacturer's recommendations as described in (Therizols et al., 2014). Single transfected cells (96) were sorted based on GFP expression from pX458 after 48h and cultured for seven days in a 96 well plate. DNA extraction and genotyping were performed 7 days after sorting using overnight incubation at 55°C with lysis buffer (10mM Tris pH 7.5, 10mM EDTA, 10mM NaCl, 0.5% SDS, 1mg/ml ProteinaseK) followed by ethanol precipitation and washes. Genomic DNA was amplified with the following primers:

SBE6.1 Fw: TTTTGGAAGCTTAAATGCCCAT
SBE6.1 Rv: CCACCACAAGCACATTCAT

SBE6.2 Fw: GCCTCCATGAAGTCCAATGG
SBE6.2 Rv: CCACCCTTGCTACTCAGGAA

Amplification was done using DreamTaq Green PCR Master Mix (ThermoFisher K1081) following the manufacturer's protocol and PCR products were assessed by agarose gel electrophoresis. Amplified products were later sequenced to further confirm homozygote deletions.

SBE6.1 Downstream Fw	CACCGgtcacattctccgtcattcag
SBE6.1 Downstream Rv	AAACctgaatgacggagaatgtgacC
SBE6.1 Upstream Fw	CACCGgtaggccagattacttgcaag
SBE6.1 Upstream Rv	AAACcttgcaagtaatctggcctacC
SBE6.2 Downstream Fw	CACCGgtctgtttgtagagaccttac
SBE6.2 Downstream Rv	AAACgtaaggtctctacaaacagacC
SBE6.2 Upstream Fw	CACCGgttagacactccagttttgtc
SBE6.2 Upstream Rv	AAACgacaaaactggagtgtctaacC

Table 20 gRNA targeting primers for deletions

8.8 Crosslinked Chromatin Immunoprecipitation

Harvested ESCs (5 - 15x10^6) were fixed in media by the addition of an equal volume of 2% methanol-free formaldehyde (Thermo Scientific Pierce PN28906; final concentration of 1%) and incubation at RT for 10 min followed by 5 min incubation with 125mM glycine at room temperature. Cells were washed in PBS. All buffers were supplemented with the following additives just prior to use (0.2 mM PMSF, 1 mM DTT, 1x Protease inhibitors (Calbiochem, 539134-1SET) and 1x phosphatase inhibitors (Roche, PhosSTOP, 04906837001)). Cell

pellets were resuspended in lysis buffer 1 (50mM Tris-HCl pH 8.1, 10mM EDTA and 20% SDS) and incubated for 10 min at 4°C. Lysates were diluted 1:10 in ChIP dilution buffer (0.1% Triton X-100, 2mM EDTA, 150mM NaCl, 20mM and Tris-HCl pH8.1) and sonicated using a bioruptor (Diagenode) at 4°C for 15 cycles of 30 sec (30 sec pause between pulses) set to high. The sonicated extract was centrifuged at 16000*g* for 10 min at 4°C and the supernatant transferred to a fresh tube and supplemented with BSA and triton X-100 to final concentrations of 25μg/ml and 1% respectively. 5% of the chromatin was retained as an input reference and the remainder incubated overnight at 4ºC on a rotating wheel with CTCF Antibody (Millipore 07-729) pre-coupled to protein G or Sheep Anti-Rat IgG Dynabeads respectively (Life Technologies; 10004D and 11035 respectively; 10μgs of antibody bound to 40μl of stock dynabeads per IP). Bead-associated immune complexes were washed sequentially with wash buffers A, B and C each for 2 x 5 mins at 4ºC on a rotating wheel followed by 2 washes in TE buffer at RT (wash buffer A - 1% Triton X-100, 2mM EDTA, 150mM NaCl, 20mM and Tris-HCl pH8.1; wash buffer B - 1% Triton X-100, 0.1% Sodium-Deoxycolate, 0.1% SDS, 1mM EDTA, 500mM NaCl, 20mM and Tris-HCl pH8.1; wash buffer C – 1% NP40, 0.1% Sodium-Deoxycolate, 1mM EDTA, 250mM LiCl, 20mM and Tris-HCl pH8.1). Chromatin was released from the beads by incubation with elution buffer (0.1 M NaHCO3 and 1 % SDS) for 15 mins at 37 ºC followed by the addition of RNaseA and Tris pH6.8 (final concentration of 20mg/ml and 100mM respectively) and incubation at 65ºC for 2 hours following by the addition of 50μgs of proteinase K and incubation at 65ºC for 8 hours to degrade proteins and reverse the cross-links. Purified DNA was isolated using QIAquick PCR Purification Kit (Qiagen). Quantitative PCR was performed on a LightCycler480 (Roche) using the LighCycler sybr green master mix (Roche) according to manufacturer's instructions.

8.9 Native Chromatin Immunoprecipitation

$5x10^6$ nuclei were prepared and resuspended in NB-R (85mM NaCl, 5.5% sucrose, 10mM TrisHCl pH7.5, 3mM $MgCl_2$, 1.5mM $CaCl_2$, 0.2mM PMSF, 5mM sodium butyrate, 5μM Sirtinol and 1mM DTT added fresh prior to use) as previously described (Gilbert et al. 2003). Nuclei corresponding to $1x10^8$ cells were digested with 50-80 Boehringer units of MNase (Sigma) for 10 min at RT in the presence of 20 μg RNase A to obtain a chromatin ladder enriched in tri-, tetra-, and some pentanucleosomes. The reaction was stopped by

adding equal volume of Stop Buffer (215 mM NaCl, 10 mM TrisHCl pH 8, 20 mM EDTA, 5.5 % Sucrose, 2 % Triton X-100, 0.2 mM PMSF 1 mM DTT and complete protease inhibitor cocktail) and incubated on ice overnight. Between 50-150 µg released chromatin were precleared with Protein G Sepharose (GE Healthcare) for 2 hr and mixed with 10 µg prebound H3K27me3 antibody (Millipore 07-449), H3K27ac antibody (Millipore 07-360) or H3K4me1 (Abcam ab8895) in the presence of 100 µg BSA and incubated for 3 hr at 4 °C. Beads were then washed 3x with Wash Buffer (150 mM NaCl, 10 mM TrisHCl pH 8, 2 mM EDTA, 1% NP40, 1 % Sodium deoxycholate, 0.2mM PMSF, 1mM DTT and protease inhibitor cocktail) and once in TE. Bound complexes were eluted with 0.1 M NaHCO3, 1 % SDS at RT. Immunoprecipitated and input DNA were purified with Proteinase K (Genaxxon) and QIAGEN PCR purification kit.

For RNA on chip, chromatin from nuclei was first prepared as for Native Chromatin Immunoprecipitation without RNase A and RNA was then extracted using TRIzol® Reagent (ThermoFisher 15596026) following manufacturer's protocol. 250 ng of each sample RNA was amplifies using WTA2 Complete Whole Transcriptome Amplification kit (Sigma) following manufacturer's protocol.

8.10 Nimblegen ChIP-on-chip from Figure Ch2-2 & Ch5-7

Ten nanograms (optimal) of input or ChIP DNA were amplified using the WGA2 whole genome amplification kit (Sigma). Amplified material was labelled with Cy3 or Cy5 by random priming according to the NimbleGen ChIP-chip protocol (Roche). In total, 2 or 3 biological replicates with dye swaps were hybridized for 20 h and washed according to manufacturer´s protocol. A custom 3x720K mouse tiling array (NimbleGen, Roche) containing 179,493 unique probes from different genomic regions, with each probe represented by 4 replicates was used. Arrays were scanned on a NimbleGen MS 200 Microarray scanner (Roche) using 100% laser power and 2 µm resolution. Raw signal intensities were quantified from TIFF images using MS 200 Data Collection software.

Arrays were scanned on a NimbleGen MS 200 Microarray scanner (Roche) using 100% laser power and 2µm resolution. Raw signal intensities were quantified from TIFF images using MS 200 Data Collection software.

Microarray data were analysed in R using the bioconductor packages Beadarray and Limma according to the Epigenesys NimbleGen ChIP-on-chip protocol 43. Scale normalization was used within replicates, to control inter-array variability. Each condition was represented by two biological replicates hybridised as dye swap experiments and enrichment scores are defined as log2 ChIP/Input signal.

8.11 Agilent ChIP-on-chip from Figure Ch5-8 & Ch5-9

One microgram of input or ChIP DNA were amplified using the WGA2 whole genome amplification kit (Sigma). All the amplified material was labelled with Cy3 or Cy5 by random priming using SureTag DNA Labelling kit according to the Agilent ChIP-chip protocol. Samples were hybridized and washed according to manufacturer´s protocol. A custom 8x60K mouse tiling array (Agilent Technologies) containing 51,504 probes was used. Arrays were scanned on a NimbleGen MS 200 Microarray scanner (Roche) using 100% laser power and 2 μm resolution. Raw signal intensities were quantified from TIFF images using MS 200 Data Collection software and extracted using Agilent Feature extraction software.

For nuclear RNA on chip, median foreground and background signal intensities from microararys were read into R using the package Limma (Ritchie et al., 2015). Signals were background corrected using normexp method and then quantile normalized to remove techincal variation. Only probes whose signal was 10% above 95 percentile of negative control probes in at least 1 array were kept for further analysis. ChIP Microarray data were analysed in R using the bioconductor packages Beadarray and Limma according to the Epigenesys NimbleGen ChIP-on-chip protocol 43. Scale normalization was used within replicates, to control inter-array variability. Enrichment scores are defined as log2 ChIP/Input signal.

	60K 58718 probes	
En2_Nom1	chr5:28,466,879-29,826,623	1,359,745
sox9	chr11:112,632,613-112,652,202	19590
kif18b	chr11:102,764,914-102,788,292	23379
nog	chr11:89,158,913-89,169,687	10,775
HoxD	chr2:74,349,931-74,723,496	373,566
ptch1	chr13:63,595,947-63,679,182	83,236
Sox4	chr13:29,038,168-29,052,795	14,628
Irf4	chr13:30,838,525-30,859,042	20,518
Ubb/Bmp4	chr14:46,679,104-46,722,760	43657
	chr14:46,997,705-47,020,448	2274
Dbx2	chr15:95,451,886-95,487,336	35451
Rad21	chr15:51,782,821-51,825,176	42356
Myc	chr15:61,814,344-61,826,001	11658
Polr2f	chr15:78,969,037-78,988,272	19,236
Lmf2	chr15:89,181,284-89,210,640	29357
Olig2	chr16:91,206,017-91,237,020	31,004
Six3	chr17:85,992,748-86,055,142	62,395
Oct4	chr17:35,634,951-35,649,171	14,221
Smc3	chr19:53,670,555-53,730,371	59,817
Pax6	chr2:105321006-105554989	233983
Foxa2	chr2:147,845,839-147,881,029	35,191
Lhx6	chr2:35,921,957-35,966,732	44,776
Nkx2-2	chr2:146,996,480-147,025,736	29,257
Nes	chr3:87,771,649-87,788,815	17,167
Sox2	chr3:34331150-34581750	250600
Klf4	chr4:55,523,676-55,550,208	26,533
Nkx6-1	chr5:102,064,277-102,101,488	37,212
Actb	chr5:143659451-143709461	50010
Nanog	chr6:122576263-122712491	136228
Gapdh	chr6:125094615-125145023	50408

Sox6	chr7:122,943,978-123,188,613	244,636
Fgf3/4/15	chr7:152,006,254-152,088,434	82,181
Nkx6-2	chr7:146,760,873-146,783,679	22,807
H19/igf2r	chr7:149758545-149858792	100247
Klf1	chr8:87,423,681-87,429,301	5,621
Sox3	chrX:58,141,136-58,153,499	12,364
Gata1	chrX:7,533,544-7,556,286	22,743

Table 21 Agilent 60K Chip-chip tiling array, genome assembly mm9

8.12 TALE design and assembly

TALE DNA binding domains specific to the promoters of Shh, En2, Rbm33, Rnf32 and Nom1 and SBE6.1, SBE2, NE sequences were assembled following the methods described in (Ding et al., 2013). Briefly, DNA binding domains specific for 15 nucleotide sequences were generated by the modular assembly of 4 pre-assembled multimeric TALE repeat modules (three 4-mer and one 3-mer) into a modified TALEN backbone in which the BamHI-BsrGI fragment containing hFokI2-2A-eGFP was replaced by a gBlocks® (IDT) fragment encoding VP64-2A-eGFP. The BamHI-BglII fragment containing VP64 of the Tale-Vp64 plasmid was deleted to generate Tale-Δ (Figure Ch8-1).

The BamHI-NheI fragment containing VP64 was replaced by double strand oligonucleotide encoding the Belmont peptide DELQPASIDP peptide (A. E. Carpenter et al, 2005) to generate Tale-BP. The BamHI-NheI cloning site was further used to fuse the Tale with the SA domain of LDB1 and the VWA domain of Med25 from gBlocks® (IDT) and CTCF, SIX3 and PARP1 from GeneArt® (Life Technologies).

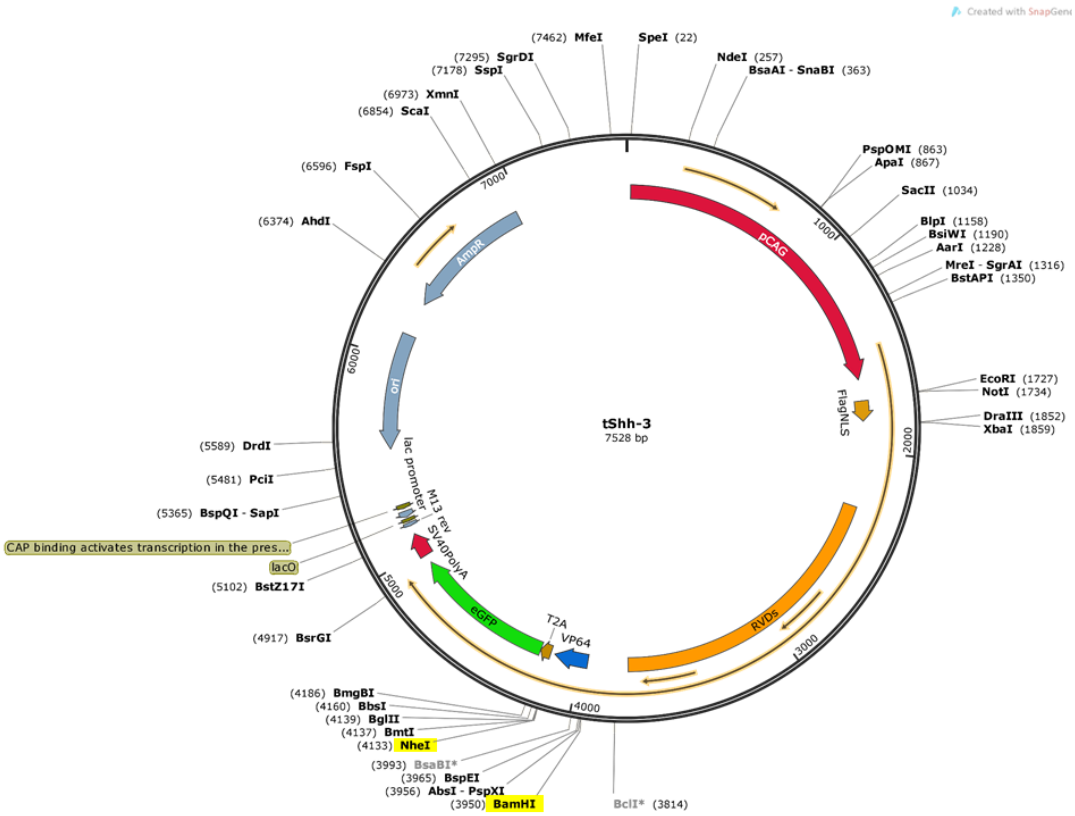

Figure Ch8- 1 tShh-Vp64 construct map.

Exemple of construct used for tShh-Vp64. RVDs sequence recognize 15 nucleotide. The Tale fusion can be cut/paste using BamH1 and Nhe1 sites and Tale-Delta can be created using BamH1 and BglII that are compatible.

8.13 Fosmid probes and Nick translation

Fosmid were minipreped after growing overnight at 37°C in L-broth containing chloramphenocol (12.5 µg/ml). Bacteria pellets were resuspended in 200 µl GTE buffer (50 mM glucose, 25 mM Tris pH8, 10 mM EDTA) containing lyzozyme and mixed vigorously, then incubated for 5 minutes at room temperature before adding 400 µl lysis buffer (0.2 M NaOH, 1% SDS) mixed by inversion and incubated again on ice for 5 minutes. The reaction was then stopped by adding 300 µl acetate buffer (5M potassium acetate 60 ml, Acetic acid 11,5 ml and Distilled water 28,5 ml) and incubating for 5 minutes on ice. Following centrifugation at 12000*g* for 5 minutes at 4°C, fosmids DNA were recovered by phenol:chloroform extraction and resuspended in 25 µl TE. RNase A (1µg) was added and incubated for 5 minutes at 37°C.

Fosmid probes were labelled, Nick translation, as followed, 2 µl Nick translation salts (0.5M Tris pH7.5, 0.1 M MgSO4, 1 mM DT, 0.5 mg/ml BSA fraction V (Sigma)) + 2.5 µl 0.5 mM dATP + 2.5 µl 0.5 mM dCTP + 2.5 µl 0.5 mM dGTP + 2.5 µl bio-16-dUTP (Roche) or 1.5 µl digoxygenin-11-dUTP + 1 µl 0.5 mM dTTP or Directly labelled with green-dUTP (Abbott Molecular 02N32-050, 00884999002913) or red-dUTP (ChromaTide Alexa Fluor 594-5-dUTP C11400) + 6 µl miniprep DNA (0.5-1 µg) + 1 µl DNase I (1:500) (Invitrogen) + 1 µl DNA polymerase I (Invitrogen). The mix was incubated for 90 minutes at 16°C. The reaction was stopped by adding 3 µl EDTA and 2µl 20% SDS, then 65 µl TE was added before purifying through a Quick spin column (Pharmacia).

Fosmid Probes

		Coordinates			Whitehead (Sanger)	
target	**chr**	**start**	**end**	**length (bp)**	**fosmid ID**	**clone**
Dpp6	chr5	27932527	27975636	43109	G135P600264D6	WI1-1085J14
CTRL	chr5	28411516	28449628	38112	G135P60453C9	WI1-2806K8
Shh	chr5	28754458	28795879	41421	G135P64333A4	WI1-574O18
SBE6	chr5	28887686	28924744	37059	G135P67311F4	WI1-442E17
SBE4	chr5	29107140	29147593	40453	G135P600205H10	WI1-2751A06

SBE2/3	chr5	29195832	29239355	43523	G135P603171G8	WI1-1275C09
ZRS	chr5	29611727	29653695	41968	G135P600929F6	WI1-1047E14

Table 22 Fosmid probes used for 3D-FISH.

8.14 2D-FISH

10-15x10^6 cells were swollen in 0.5% trisodium citrate/0.25% KCl followed by fixation and 3 washes in methanol acetic acid (MAA – 3:1 vol/vol). Slides were incubated in 100 µg/ml RNase A in 2 x SCC for 1 hour, washed in 2 x SCC and dehydrated through an alcohol series. Slides were denatured in 70% formamide/2 x SCC for 75 s at 70°C. Between 80-120 ng of biotin- and digoxigenin-labeled fosmid probes were used per slide, with 8-12 µg of mouse Cot1 DNA (Invitrogen) and 10 µg salmon sperm DNA. Probes were denatured at 70°C for 5 minutes, reannealed with Cot1 DNA for 15 minutes at 37°C and hybridized to the denatured slides overnight at 37°C. Slides were washed 4 x 3 minutes in 2X SSC at 45°C, 4 x 3 minutes in 0.1X SSC at 60°C and transferred to 4X SCC, 0.1% Tween 20. Slides were counterstained in 0.5 µg/ml DAPI. The BAC used to highlight Shh regulatory region is RP24-323C22 coordinates chr5:28,832,721-29,003,817 in mm9

8.15 3D-FISH

1x10^6 ESCs or NPCs were seeded on slides for 5h. Cells were fixed in 4% PFA for 10 minutes at room temperature and then permeabilized using 0,5% Triton X for 10 minutes. Approximately 150 ng of labelled fosmid probes were used per slide, together with 15 µg of mouse Cot1 DNA (GIBCO BRL) and 10 µg salmon sperm DNA. Probes were denatured at 70°C for 5 min, reannealed with CotI DNA for 15 min at 37°C and hybridized to the denatured slides overnight. DNA was denatured at 80°C for 20 minutes. For dig- or biotin-labelled probes, the last 4X SCC, 0.1% Tween 20 incubation was followed by an extra 5 minutes incubation with blocking buffer (4X SCC, 5% Marvel) and biotinylated probes were detected using fluorochrome-conjugated avidin (FITC or Texas Red) (Vector Laboratories) then biotinylated antiavidin (Vector Laboratories) followed by fluorochrome-conjugated

avidin. Digoxigenin-labeled probes were detected using FITC-conjugated antidigoxigenin (Vector Laboratories) followed by FITCconjugated anti-sheep (Vector Laboratories).

8.16 Image Capture and Analysis

For 2D FISH, slides were imaged using a Hamamatsu Orca AG CCD camera (Hamamatsu Photonics (UK) Ltd, Welwyn Garden City, UK) and a Zeiss Axioplan II fluorescence microscope with Plan-neofluar objectives, a 100W Hg source (Carl Zeiss, Welwyn Garden City, UK) and Chroma #83000 triple band pass filter set (Chroma Technology Corp., Rockingham, VT) with the excitation filters installed in a motorised filter wheel (Prior Scientific Instruments, Cambridge, UK).

3D-FISH slides were first imaged with a Hamamatsu Orca AG CCD camera (Hamamatsu Photonics), Zeiss Axioplan II fluorescence microscope with Plan Apochromat objectives, a Lumen 200 W metal halide light source (Prior Scientific Instruments) and Chroma #89014ET single excitation and emission filters (Chroma Technology Corp.). Images were deconvolved using a calculated PSF with the constrained iterative algorithm of Volocity (PerkinElmer).

Super-resolution images were acquired using structured illumination microscopy following published protocol (Gustafsson et al., 2008). Samples were prepared on high precision cover-glass (Zeiss, Germany). 3D SIM images were acquired on a N-SIM (Nikon Instruments, UK) using a 100x 1.49NA lens and refractive index matched immersion oil (Nikon Instruments). Samples were imaged using a Nikon Plan Apo TIRF objective (NA 1.49, oil immersion) and an Andor DU-897X- 5254 camera using 405, 488, 561 and 640nm laser lines. Z-step size for Z stacks was set to 0.120 um as required by manufacturers software. For each focal plane, 15 images (5 phases, 3 angles) were captured with the NIS-Elements software. SIM image processing, reconstruction and analysis were carried out using the N-SIM module of the NIS-Element Advanced Research software. Images were reconstructed using NiS Elements software (Nikon Instruments) from a z stack comprising of no less than 1um of optical sections. In all SIM image reconstructions the Wiener and Apodization filter parameters were kept constant.

Image analysis was carried out using the Quantitation module of Volocity (PerkinElmer). Reconstructed SIM data was directly uploaded an analyzed on Volocity. The statistical significance of differences in mean-squared interprobe distances was assessed using the nonparametric Mann-Whitney U test to examine the null hypothesis. Each data set consisted of 20 to 50 nuclei (40 to 100 loci). Biological replicates are shown under their p-values.

The data analysis is done using Volocity, widefield images were deconvolved using point spread function (PSF) set up for the 3D-widefield microscope. Reconstructed SIM data was directly uploaded an analyzed on Volocity.

8.17 Single Cell qRT-PCR

RNA reverse transcription and cDNA pre-amplification from single cells were performed as previously described (Dalerba et al. 2011) with some modifications. Each well of a 96-well PCR plate was loaded with 5 µl 2X Reaction Mix, 0.2 µl Superscript III RT/Platinum Taq Mix (with RNaseOUT Ribonuclease Inhibitor) (Invitrogen Cells Direct One-Step qRT-PCR kit, Life Technologies), 2.5 µl primer mix (containing 200 nM of each gene-specific primer), 1.3 µl H2O. Single-cell suspensions from transfection or NPC differentiation were sorted using their GFP reporter into separate wells of the 96-well PCR plate. 32 cells were sorted into one well, to be used for serial dilution for generation of qRT-PCR standard curves. RNA reverse transcription and 22 cycles of cDNA pre-amplification were performed as previously described (Dalerba et al. 2011). The cDNA was diluted 1:5 in H2O and qRT-PCR were performed as above using 9 µl of this diluted cDNA.

8.18 Chromatin nicking with bleomycin

$1x10^6$ Cells in six-well plates (for DNA extraction) or on slides (for FISH analysis) were treated with bleomycin (300 µM) for 10 min (Naughton et al., 2013). Slides were then washed and fixed with 4%PFA to proceed to the FISH. For DNA extraction cells were lysed in 150 mM NaCl, 1% SDS, 10 mM EDTA and incubated for 1 h at 55 °C with 100 µg ml^{-1} proteinase K. DNA was purified by phenol-chloroform extraction and recovered by ethanol precipitation. DNA was suspended in 50 mM NaOH, 1 mM EDTA, 3% Ficoll, 0.02% bromocresol green and 0.04% xylene cyanol and fractionated on a 0.7% agarose gel

in 50 mM NaOH, 1 mM EDTA at 2 V cm^{-1} with buffer circulation for 20 h. The gel was stained with ethidium bromide in 1× TAE and imaged on a FLA5100 laser scanner (Fuji).

8.19 3C library preparation

Cells were fixed with 1% formaldehyde for 10 min at room temperature. Cross-linking was stopped with 125 mM glycine for 5 min at room temperature followed by 15 min on ice. Cells were centrifuged at 400*g* for 10 min at 4°C, supernatants were removed, and cell pellets were flash-frozen on dry ice.

Cell pellets were treated as previously described (Dostie and Dekker, 2007; Ferraiuolo et al. 2010). Briefly, 10-15x10^6 fixed cells were incubated for 15 min on ice in 200 µL of lysis buffer (10 mM Tris at pH 8.0, 10 mM NaCl, 0.2% NP40, supplemented with fresh protease inhibitor cocktail). Cells were then disrupted on ice with a dounce homogenizer (pestle B; 2 × 20 strokes); cell suspensions were transferred to Eppendorf tubes and centrifuged at 2000*g* for 5 min. Supernatants were removed, the cell pellets were washed twice with 100 µL of 1× EcoRI buffer (New England Biolabs), and the cell pellet was resuspended in 100 µL of 1× EcoRI buffer and divided into two Eppendorf tubes. We added 1× EcoRI buffer (337 µL) to each tube, and the mixture was incubated for 10 min at 65°C with 0.1% SDS. Forty-four microliters of 10% Triton X-100 was added before overnight digestion with 400 U of HindIII. The restriction enzyme was then inactivated by adding 86 µL of 10% SDS and incubation for 30 min at 65°C. Samples were then individually diluted into 7.62 mL of ligation mix (750 µL of 10% Triton X-100, 750 µL of 10× ligation buffer, 80 µL of 10 mg/mL of BSA, 80 µL of 100 mM ATP, 3000 cohesive end units of T4 DNA ligase) and incubated at for 2 h 16°C.

3C libraries were incubated overnight at 65°C with 50 µL of Proteinase K (10 mg/mL) and an additional 50 µL of Proteinase K the following day for 2 h. The DNA was purified by one phenol and one phenol–chloroform extraction and precipitated with 0.1 vol (800 µL) of 3 M NaOAc (pH 5.2) and 2.5 vol of cold EtOH (20 mL). After at least 1 h at −80°C, the DNA was centrifuged at 20,000*g* for 25 min at 4°C, and the pellets were washed with cold 70% EtOH. DNA was resuspended in 400 µL of TE (pH 8.0) and transferred to Eppendorf tubes for another phenol–chloroform extraction and precipitation with 40 µL of 3 M NaOAc (pH 5.2) and 1.1 mL of cold EtOH. DNA was recovered by centrifugation and washed eight

times with cold 70% EtOH. Pellets were then dissolved in 100 µL of TE (pH 8.0) and incubated with1 µL of 10 mg/mL RNase A for 15 min at 37°C.

8.20 5C primer and library design

5C primers covering the *USP22* (mm9, chr11: 60,917,307–61,017,307) and *Shh* (mm9, chr5: 28317087-30005000) regions were designed using my5C.primer (Lajoie et al. 2009) and the following parameters: optimal primer length of 30 nucleotides (nt), optimal TM of 65°C, and default primer quality parameters (mer: 800, U-blast: 3, S-blasr: 50). Primers were not designed for large (>20-kb) and small (<100-bp) restriction fragments, low-complexity and repetitive sequences, or when there were sequence matches to more than one genomic target. The*USP22* regions was used to assess the success of each 5C experiment but was not used for further data normalization or quantification.

The universal A-key (CCATCTCATCCCTGCGTGTCTCCGACTCAG-[5C-specific]) and the P1-key tails ([5C-specific]-ATCACCGACTGCCCATAGAGAGG) were added to the forward and reverse 5C primers, respectively. Reverse 5C primers were phosphorylated at their 5′ ends.

8.21 5C library preparation

5C libraries were prepared and amplified with the A-key and P1-key primers as described previously (Fraser et al. 2012). Briefly, 3C libraries were first titrated by PCR for quality control (single band, absence of primer dimers, etc.) and to verify that contacts were amplified at frequencies similar to that usually obtained from comparable libraries (same DNA amount from the same species and karyotype) (Dostie and Dekker 2007, Fraser et al. 2010). In general, we used 1–11 µg of 3C library per 5C ligation reaction.

5C primer stocks (20 µM) were diluted individually in water on ice and mixed to a final concentration of 0.002 µM. Mixed diluted primers (1.7 µL) were combined with 1 µL of annealing buffer (10× NEBuffer 4 [New England Biolabs]) on ice in reaction tubes. Salmon testis DNA (1.5 µg) was added to each tube, followed by the 3C libraries and water to a final volume of 10 µL. Samples were denatured for 5 min at 95°C and annealed for 16 h at 48°C. Ligation with 10 U of Taq DNA ligase was performed for 1 h at 48°C. One-tenth (3 µL) of

each ligation was then PCR-amplified individually with primers against the A-key and P1-key primer tails. We used 26 or 28 cycles based on dilution series showing linear PCR amplification within that cycle range. The products from two to four PCR reactions were pooled before purifying the DNA on MinElute columns (Qiagen).

5C libraries were quantified on agarose gels and diluted to 0.0534 ng/µL (for Xpress template kit version 2.0) or 0.0216 ng/µL (for ion PGM template OT2 200 kit). One microliter of diluted 5C library was used for sequencing with an ion PGM sequencer. Samples were sequenced onto ion 318 chips following either the ion Xpress template kit version 2.0 and ion sequencing kit version 2.0 protocols or the ion PGM template OT2 200 kit and ion PGM sequencing 200 kit version 2.0 protocols as recommended by the manufacturer (Life Technologies).

8.22 5C data analysis

Analysis of the 5C sequencing data was performed as described earlier (Berlivet et al. 2013). The sequencing data were processed through a Torrent 5C data transformation pipeline on Galaxy (https://main.g2.bx.psu.edu). Data were normalized by dividing the number of reads of each 5C contact by the total number of reads from the corresponding sequence run. All scales correspond to this ratio multiplied by 10^3.

Name	Number of reads	Number of Interactions	Number of used normalized reads
ESC biological replicate 1	45024284	**116,281**	**1,950**
ESC biological replicate 2	42947512	**116,281**	1,924
NPC biological replicate 1-technical replicate 1	92081330		
NPC biological replicate 1-technical replicate 2	2079276		
NPC biological replicate 1-Average		**116,281**	**1,879**
NPC biological replicate 2-technical replicate 1	3984330		
NPC biological replicate 2-technical replicate 2	20101490		
NPC biological replicate 2-Average		**116,281**	**1,877**

tSBE6-Vp64 biological replicate 1	90452930	**116,281**	**1,954**
tSBE2-Vp64 biological replicate 1	1.01E+08	**116,281**	**1,909**
tSBE6+tSBE2-Vp64 biological replicate 1-technical replicate 1	68323948		
tSBE6+tSBE2-Vp64 biological replicate 1-technical replicate 2	79792180		
tSBE6+tSBE2-Vp64 biological replicate 1 Average		**116,281**	**1,926**
tSBE6+tSBE2-Vp64 biological replicate 2-technical replicate 1	19389432		
tSBE6+tSBE2-Vp64 biological replicate 2-technical replicate 2	24309904		
tSBE6+tSBE2-Vp64 biological replicate 2 Average		**116,281**	**1,864**

Table 23 5C sequencing technical and biological replicates reads.

Appendix

Predicted putative sites in human SBE6.1

Model ID	Model name	Score	Relative score	predicted site sequence
MA0761.1	ETV1	5.396	0.81765789	AGCGTATGTG
MA0761.1	ETV1	5.299	0.81582659	GAAGGAAATA
MA0882.1	DLX6	6.97	0.89451043	CCAATTTT
MA0722.1	VAX1	4.727	0.81707625	CCAATTTT
MA0747.1	SP8	7.728	0.82568109	CCCAGCCCCACC
MA0722.1	VAX1	4.376	0.80711244	TTCAATAC
MA0747.1	SP8	6.661	0.80585165	GACATCCCCCGT
MA0761.1	ETV1	6.509	0.8386707	GCCGGACACA
MA0761.1	ETV1	6.995	0.8478461	TGCGGAAGGG
MA0747.1	SP8	9.489	0.85840802	ACAACGCCCTCC
MA0722.1	VAX1	4.936	0.82300912	CAAATAAC
MA0882.1	DLX6	3.671	0.80896883	GCAAATAA
MA0761.1	ETV1	8.063	0.86800933	AACGGAAGGT
MA0761.1	ETV1	4.485	0.80045873	GACGGAAAAC
MA0747.1	SP8	6.4	0.80100116	GCTCCTCCTCCT
MA0722.1	VAX1	5.21	0.83078713	CTCATGGC
MA0722.1	VAX1	6.426	0.86530564	CAAATGAT
MA0747.1	SP8	6.635	0.80536846	ACAAAGCCCGCT

MA0631.1	Six3	12.164	0.83658066	AAAAGGGGATCATAAGC
MA0722.1	VAX1	5.308	0.83356905	CTTATGAT

Predicted putative sites in human SBE6.2

Model ID	Model name	Score	Relative score	predicted site sequence
MA0747.1	SP8	7.953	0.82986255	ACCCCACCCAGA
MA0747.1	SP8	7.145	0.81484645	CCCACCCCACCC
MA0747.1	SP8	8.476	0.83958214	GCCCACCCCACC
MA0747.1	SP8	6.641	0.80547997	GCCACCTCTCCT
MA0882.1	DLX6	6.926	0.89336953	GCAATTTC
MA0722.1	VAX1	4.386	0.80739631	CCCATGAT
MA0882.1	DLX6	3.763	0.81135435	GACATTAT
MA0882.1	DLX6	5.218	0.84908185	CCAATAAT
MA0722.1	VAX1	5.064	0.82664264	CCAATAAT
MA0722.1	VAX1	5	0.82482588	CCAATGGT
MA0722.1	VAX1	4.873	0.82122074	CAAATGGC
MA0882.1	DLX6	6.534	0.88320515	GCAATTTT
MA0882.1	DLX6	5.186	0.84825211	ACAATGAA
MA0722.1	VAX1	6.058	0.85485925	TTCATTGT
MA0722.1	VAX1	4.537	0.81168273	CTCAATAC

MA0882.1	DLX6	3.514	0.80489789	CTCATCAC
MA0722.1	VAX1	4.642	0.81466336	CTCATCAC
MA0747.1	SP8	8.445	0.83900602	CCCCAGCCCACA
MA0761.1	ETV1	4.574	0.802139	AACGGCAGCA
MA0747.1	SP8	7.937	0.82956521	CCCACACCCTCA
MA0747.1	SP8	6.723	0.80700388	AACCCTCCCCAA
MA0747.1	SP8	6.402	0.80103832	ACCAACCCTCCC
MA0761.1	ETV1	6.107	0.83108117	CCCGGATGAC
MA0761.1	ETV1	9.018	0.88603919	TGCGGAAGTC

Table S 1 Jaspar scores for forebrain TFBS in SBE6.1 and SBE6.2 in human.

	Type	Genomic sequence (5' to 3')	HindIII position Start	HindIII position End
Shh (chr.5; mm9)				
3	R	CTTCCAGTATTGGGTTACAGTTAATGGAGT	28317087	28319149
5	F	AGCATAGAGTGTGTGTAGGTGCTGCCTAAG	28319935	28324734
7	R	CTTTCTCTCATCCCTACACTAACCAGGCCT	28325862	28331585
8	F	GGTAAGAGTCCCAAAGAACAGCTTGTTAAG	28331586	28333893
9	R	CTTGACACACCTACCCTCTAAGTAATCAAT	28333894	28336649
10	F	AACACCTCTAGCATGATAGCACTTTGCAAG	28336650	28340960
11	R	CTTAGGATGTGCCTCTACTGTGGGGG	28340961	28344675
12	F	CAAAACCCTAGAAGCCACAGGGACCAAG	28344676	28347365
13	R	CTTGCCAGTTTATCTAGGTAGCCTGCCAG	28347366	28348911
18	F	GGAGACCCACACTAAGGGCCTCAAG	28355883	28370234
19	R	CTTGGATTGGCGTGGCTCTGAGTCAT	28370235	28373716
20	F	TGTGTTTTAGGGATGAGGGATTCTTTAAAG	28373717	28374948
25	R	CTTCCTTTCTGGTATCTATTGACCTTCCCT	28385728	28388130
27	F	TTTGATAGTGCTGTTTCCTGTGGCTAGAAG	28388373	28403099
30	R	CTTCTCCTGTAAGATGGCAATTTATTTATT	28416843	28419608
31	F	TTCTAAATATAATCCAGAGAGAAGGCTAAG	28419609	28428926
32	R	CTTTGTCATTCCACCATGGTTGTGGTGAAC	28428927	28447705
34	F	CAATAAAGGTAGAACTTGGGCCCAGTGAAG	28452395	28454456
35	R	CTTGAGCTCACATATGGGACACTCTTGACA	28454457	28458325
38	F	ATGGGGCCGGATTTAACTCAACAATCAAAG	28463308	28463573
41	R	CTTCTCTAGCTAGGCCAGCATAATGTACCG	28469222	28469868
42	F	AAAAATAAAAAGGAGGCCTGGATTCTGAAG	28469869	28470959
44	R	CTTAGCTCAAAATGTAGGAAATGGCCATTC	28476584	28484052
45	F	AAAATCTCCCTGGAGTCAAAGGGTTAAAAG	28484053	28486216
46	R	CTTGTGAACAGTCCCACCAGGTCACTG	28486217	28491939
47	F	CCACCCCCAGTATCTGCAACCTCAAG	28491940	28496579

49	R	CTTAAGGTGGGGGTGACACAGTCCAAAG	28504618	28506235
50	F	GCAAGAGCCCACCAGGGTCAAG	28506236	28509301
52	R	CTTCAGTTTGGGTGACACATGCAGGACAAA	28509323	28510999
53	F	GTGATGTCTCCCCTGTGAGCAGGAAG	28511000	28511814
56	R	CTTATCCTCTTCTGTGTCTAGTTGAAGTGG	28511842	28514356
57	F	TTCTCATGTCAAGATCCACATAATATCAAG	28514357	28515715
58	R	CTTGTTGAGCTCATGTAAATGCCTATGGAT	28515716	28528247
59	F	TTACATATCAGCTGCTGTATCCATCACAAG	28528248	28532168
61	R	CTTGAAAAGCACAGATAAAAATGCCATTTG	28534318	28537153
62	F	CACAGGGCTCTTTCATAGCCTAAGAACAAG	28537154	28541461
63	R	CTTCCTTGGGCAAGTGGTATCTTCCTTAGC	28541462	28542832
64	F	CACAGTCACCCCCTGTACCCCAAAG	28542833	28543234
65	R	CTTTGACAAAGTGATGCATCTAAGATCCTA	28543235	28543365
68	F	GGTACTTCTACAGTGGGGGAGGATGTAAAG	28547549	28547729
69	R	CTTCTCATATGACTCTGGTTTCTTGGCCCA	28547730	28548357
73	F	TCTATGTATAAGCCACACCAAGGAAAGAAG	28559901	28566954
74	R	CTTAGCATGGGACTCAGAAAACAAAATAGG	28566955	28570463
75	F	TGGCAAATCAGAAAAACTCTTTGGATTAAG	28570464	28571000
76	R	CTTGAAATTGAAGTATCTCTCTCAGCACCT	28571001	28572084
77	F	GAGTTCAAGAGCCCCCAAATCCCTCTAAG	28572085	28576299
78	R	CTTCCAGAAGATCTGCAGCAACTCTCTCTC	28576300	28577015
79	F	GGAGGCAGAGCCTCTGAGTCACAAG	28577016	28580658
80	R	CTTGAAGCATGTGTGGACTCCCATTCTTCT	28580659	28581218
81	F	CTCCAGACTGAGACCTTCCTGAGACCAAAG	28581219	28585322
83	R	CTTGCGGTTGTAGCTAAGAGTGAATTTGAA	28585390	28593411
84	F	GGAAAGGTACTCTGGGGTGCATCACAAAG	28593412	28595934
86	R	CTTTGGGGTGGGAACAAGGAGACTTCAC	28604862	28605362
90	F	GACCTGACACCTTTGGGATGAAAGTGTAAG	28611343	28619284
92	R	CTTGACAGAGGAGCCTAAAAGGTGACTTAA	28619329	28620319
93	F	ATTGGCTATGTAGATGAAGATGGTCCCAAG	28620320	28627997
94	R	CTTTCATTGAGACACTCCCTCAGCCTCAGT	28627998	28631580

95	F	CGTGTTACAGTTAGCTACACCCTCAAAAAG	28631581	28635856
96	R	CTTAGCAGTTTCTGTAAAAAATAAAAGTAC	28635857	28638808
97	F	GTAGCGTTCGCGCGCCTCAAG	28638809	28643392
98	R	CTTGTGCAGTACTAAATCATAATGCCATAA	28643393	28646446
99	F	TAGCATAGGGGTTATGGATGGACTCAAAAG	28646447	28647705
100	R	CTTTAAAAGGTACAATGATAGAAGAAATAG	28647706	28650692
101	F	TGAGTTTTCAGTAACCACTATAAAAAGAAG	28650693	28651422
102	R	CTTCCTATGGCTGAGAACTGCTTAGATAAT	28651423	28652423
103	F	GCTGATTCCTTTGCTGACTGGAGTGTGAAG	28652424	28656081
106	R	CTTTTGGTTGTTTCAACCATTTTTCACTTA	28662072	28668894
107	F	GTCTTCTTATCCCTGGGTAAATTGTTAAAG	28668895	28669869
109	R	CTTCCTCTCTCAGGAAACCAGTCTTCTGAG	28676542	28681826
111	F	CAGCATGGCTGTGAGGGAAAGTAGCTAAG	28681874	28682959
112	R	CTTCCAACAGTACATTATCCTAAGCGTCTA	28682960	28684009
114	F	CAATTTCAGTGCCAGCCTCTCTCGGTAAG	28685148	28687114
115	R	CTTACACCAATTAACTTCTTAGAAGTAGAC	28687115	28688407
117	F	CCTGTTTGCACTGTGTCTTCTCACAGGAAG	28689808	28693301
118	R	CTTAAATCACGAAGTACTGAGGCTTACCAA	28693302	28698428
119	F	AAAAGAAAAAGAATCACCATGACTCTTAAG	28698429	28705774
120	R	CTTATTCACAGGCCATTCTGGTAGGAACAT	28705775	28710085
121	F	CATGGTAACATCGTGTGTAGATAGAAAAAG	28710086	28710865
122	R	CTTATGTGGAGAACTAACACCATCATAAAT	28710866	28715944
123	F	AACTTTGTGGCACCTTTCTCCTCCAGAAAG	28715945	28717043
124	R	CTTTACCACTGCGGAAGGGGGAAAACA	28717044	28718937
128	F	CCTTCTGTGTGATCATCTGACACATAAAAG	28723250	28724961
129	R	CTTATACTGGGTGGAGGTCAATTCTGGACT	28724962	28731808
130	F	GTTGCAGACTGAGGGGCTCCAGAAG	28731809	28735435
131	R	CTTTAACCTGGCTCTGCTCTCAGAATGAGG	28735436	28738817
132	F	ATTTTCTTGTGGCATTATTAGGCAGGTAAG	28738818	28743466
133	R	CTTTAATGTTCTGGTTTGTTGTTGTTCTAA	28743467	28748045
134	F	CCTGCAGTCAGGGAACCGAGAGAAG	28748046	28750565

139	R	CTTTTATGTGGCTCTGCTTTTGTATTACAA	28759100	28763978
140	F	ATATTTGGATGTTCTGTCAGTGGCCTGAAG	28763979	28767078
143	R	CTTGTTCCCCGTACCCACATAAAAGGCC	28770697	28780563
144	F	CCAGAGACCCCTCCATCTGCTCAAAG	28780564	28782634
145	R	CTTTTCCCTCACCCCATTGAAAGAAGGGAG	28782635	28789524
147	F	CATCTGATTGGCCAAGCCGCACAAG	28789547	28793709
149	R	CTTACTACCAGTCCTTTGCTCTGTCTAATA	28794058	28797606
151	R	CTTGGCTAACATTGGACAACCCAAGTGTTT	28798546	28799298
152	F	TAGAAAAGATGCTGGGAACCTCATTCTAAG	28799299	28801546
154	R	CTTCTGGACACCCAGAAATGTGCGTCTC	28805019	28805807
155	F	GTGGAGCCATCATGGAAATTGCATGGAAG	28805808	28806894
158	R	CTTCTCCTAAGAACAGCTAGACCTATGCAC	28821671	28827662
162	F	AGAACCACAGGATACCCATAAGAGCCAAAG	28837190	28843289
163	R	CTTCAAAGCTGCAGTGCTTTGAAGTGTCTG	28843290	28847606
168	F	ATATCTACACCAGCTTTCTAAAATGGAAAG	28870608	28876511
169	R	CTTTATTGCCAGGTCAAATGATTTAAACAT	28876512	28878134
170	F	CTGTGATCTGAAGGTGTAAGCTGAGATAAG	28878135	28878784
171	R	CTTTGAAGGAGACCCTATTTCCTATGTGGG	28878785	28879286
173	F	GGTGTGCAGCCAGTGTGCATATTAGACAAG	28879305	28880400
174	R	CTTTATTCACCTCTGACATGCAAGCCAACA	28880401	28880839
176	F	GGCTGCAAAAGTTGGGTCTCATTTGTGAAG	28882358	28883821
177	R	CTTTGGAAGGCTGGGTGGTCAGC	28883822	28884803
178	F	CTTGGGTAAAGCTATACTGGATGCGGCAAG	28884804	28887360
179	R	CTTCCAAAATTCTATTTTGGGAAAAAATGA	28887361	28889232
180	F	TTACCTAGGTAACTGCTGCCCTTTCAGAAG	28889233	28895175
181	R	CTTCCAGAAACTGTTTACTTCCTTCTGGAG	28895176	28899338
191	F	TTGTTATTCCTGGGAGACTTAATTGGCAAG	28923745	28923931
193	R	CTTCCAGGTCACGTTAAAGATATTTCAGTA	28925837	28929215
197	F	CCACAACCTGCCCCTACGGTGTAAAG	28942022	28945936
198	R	CTTAGAAGGACCCTAGAATGGTCCCCTGAA	28945937	28949569
199	F	GGTATGGGATATCCTTTGGGGTTCACTAAG	28949570	28950647

200	R	CTTTGAATGAGGTCAATAAAAATCTACCTC	28950648	28951388
201	F	CTTCACTCCCTCAGTTACCAAGCCACAAAG	28951389	28952477
204	F	CAGGATGCCTTAGGAGACACGAGGAAG	28954090	28954407
207	R	CTTTTCAAAAAGCCACATGAAAATCTACTA	28971638	28978794
212	F	CAGGCACCTAAGTGTTAGAGAAGTTGGAAG	28997868	29000514
213	R	CTTATGAGTGCAAGGTCTGCCCAGGTTG	29000515	29003765
214	F	TACAAGTCTCATCTGAGCCCTCCAAAAAAG	29003766	29004021
215	R	CTTAGTTGTTCCTTCAGTGTCAGTTACTTC	29004022	29005070
216	F	GATGGGTCTTTCAGAGTCTGTTCCCTGAAG	29005071	29006498
217	R	CTTTTATTAAAAGGCCATGGGGCCATGGAG	29006499	29008669
218	F	TGTTCACCAAAATTTATTCTAAAAGGCAAG	29008670	29011994
219	R	CTTCTAATGAACCTGCTCCTGACCGCATG	29011995	29013866
220	F	CAAGATTACCCTGAAGTGCCGGTACAGAAG	29013867	29014066
221	R	CTTCAGGAGTCTAACTGCAATAATAATGCA	29014067	29020057
222	F	GCTCTGTGTGTCCTTCAGCTCTCTGAAG	29020058	29022523
223	R	CTTTGCTGTTGGTCAAAGGTAGCAGCTGA	29022524	29028120
224	F	AGCAGGCTTCCTCCTAGGATTATAATGAAG	29028121	29029629
225	R	CTTCGGGCCCCGGAGGGAGG	29029630	29032468
226	F	CTGAGTCTCAAGCAGCTAGCTTTCAGAAAG	29032469	29038542
227	R	CTTTTGAGAGGTGTGAGACAATCAAAATAA	29038543	29039555
230	F	CTTCCCAGGCTTTGAAGGGAACACACTAAG	29048476	29050003
231	R	CTTCAATTTGTGAGCCTCTACAAAACCTCA	29050004	29051239
232	F	GGTCCACTGGCAGCCCAAGAAG	29051240	29058423
233	R	CTTCCTTTTCATCTTGAATCAGCCTATAGA	29058424	29060961
234	F	AGTGTTTAGGGTTCTAAGGACATGGCAAAG	29060962	29068732
237	R	CTTTCTCCCAGTGATGCTGTTAGTTGTTCC	29078836	29081207
238	F	AGTTTCCTTGAAAATCATGGCCCAGCAAAG	29081208	29082327
239	R	CTTGCAATACTTCCAAGAGAAAGAGAACAC	29082328	29083134
241	F	TCACACACCTGGCAGCTGGACAAG	29084039	29096781
242	R	CTTTCAATTACTGCTCTAGCACAATGCCTG	29096782	29100883
243	F	AAGATAAAAACCTTCCATCTTAGAACAAAG	29100884	29101246

244	R	CTTGAAAAGAGGAATATTCGGGAATGGAGG	29101247	29109459
245	F	TAGGAAAAGGGACTTAATTTGTCATCAAAG	29109460	29111881
246	R	CTTAGATCTGGTGGTAGAAGGGATACTGGA	29111882	29113252
247	F	CGTATAGACTTAAATGAATTAGAACAAAAG	29113253	29117196
248	R	CTTTGCTGTCTGCTCAGAGGAGAACTCTGT	29117197	29119739
251	F	AGAAGGATTGTTTTATTCCTGTCCTTAAAG	29130489	29131877
252	R	CTTAGGCCATGAAGGAAGATGGCTTTGACA	29131878	29136802
254	F	GTTTTTTTCTGGCACAAGCTACCACTTAAG	29137043	29154044
255	R	CTTCCAATGATCCTCTCTGCTCCAATGAAA	29154045	29155707
256	F	TTTATTCTAACACTTATCCCATCCTGCAAG	29155708	29159386
258	R	CTTGGGGACCATTTAAAATATGTTCTAGAT	29159409	29160128
259	F	CTATATTTATTTAAAGTACAAAAACCTAAG	29160129	29166826
260	R	CTTGCCATGTAAGATGGGATATCTGGCCAG	29166827	29168466
261	F	GCACCGAGACCAGCTTCTGAGATCAAG	29168467	29173627
262	R	CTTATTTATCAAGTACAGTTGCTCAAGTAT	29173628	29173865
263	F	CATTAAACTGTGTTGAACCTATTTATTAAG	29173866	29174415
264	R	CTTGCAGCTCCAGACATTTACCCATCTCTG	29174416	29179395
265	F	TGGCACCTCAAATTGAGACCTTGCTTTAAG	29179396	29180591
266	R	CTTCCTCCTTACCCCGGCTGCTAATC	29180592	29184579
267	F	AGTGTTTCTATTTTTATGCTGAGTCCCAAG	29184580	29185047
268	R	CTTTGATATTACAGTGATGAATTGATATGT	29185048	29185852
269	F	TGAGAACAGTATTTTACTTAATTTGAGAAG	29185853	29188276
271	R	CTTAGACTTCCTTTTAGAACATTCTTGGTT	29188338	29191726
272	F	TTGAAAGCTGATTTCAAACAATGATTAAAG	29191727	29196352
274	R	CTTATCAACAACACCTGCATTTTAAAAGAC	29198345	29199536
275	F	AGCTATCATTTGGTTAAAAACTGTTAGAAG	29199537	29203806
276	R	CTTCATGCTGGCAGACAAAGTAAATTCGGG	29203807	29203920
278	F	GGCACATGCTGGGTCCCAGATAAG	29206157	29209241
279	R	CTTCCTTGTAAATATCTGGAATAGAAAGAG	29209242	29216468
280	F	CATGGGAGGTGTCAAACGGATTGGTGAAG	29216469	29220327
281	R	CTTGGAGATTTCCAGGTCTGCTTCTACATT	29220328	29223230

283	F	GCCACAGTCTGGAAGCACAGATCCAAG	29230618	29233018
284	R	CTTCCATCAGATCTGTGTTGTCCAAGAATC	29233019	29236380
285	F	CCAGGGTCATCTATGTGCATGCTCACAAG	29236381	29237108
287	R	CTTGCTAACACCGTTTGAGGTGGAGATGC	29237129	29240127
288	F	TGTCATCCCTTGTTCTAGTTTGACTCTAAG	29240128	29243266
289	R	CTTCTCCTGGTCATCATCTGGCCAGTATCA	29243267	29246728
290	F	GCGTAGTAAACACGGGGGTTATAAGCTAAG	29246729	29247568
291	R	CTTGGCTACACTGCATTTCCTTGTATTATA	29247569	29250429
292	F	GGGAGCAATTCTTAAGAGTGCTTTTCTAAG	29250430	29250736
295	R	CTTTGAAGATTTCCAAGAGTTCAAACCCAG	29253038	29259186
296	F	CTAGGTTTCTTGGGGAGAAGGGCTATGAAG	29259187	29262463
297	R	CTTGAAAAGTGAAGAGATATCCTACAGAAT	29262464	29264721
298	F	GGAGCAGCCAGTAGCCCCAGAAG	29264722	29268159
299	R	CTTGCTTTGCCATTGGACCCTTGTCAG	29268160	29268267
300	F	TGATGCTATCTCTCTTCAAAGGAGGAGAAG	29268268	29270862
301	R	CTTTAAAGCAACAATATGATATGCATCATC	29270863	29272035
303	F	TGGTAAGAAAATGAAATGTAAGGTACGAAG	29277402	29281527
304	R	CTTCCATCTTTTTTCATTTAGAAGCACCAG	29281528	29287941
309	F	GATGGATCTGGAAGGAGGGACACCAAAAG	29292944	29296257
313	R	CTTTAAATTCCTGCTTGTAAAAATTGTTTT	29305044	29309647
314	F	CCAGCCTGTGATGACAGGTGGTAGAAG	29309648	29310307
316	F	ACCTTTCCTCTTCCTAACTGCCATGCAAAG	29310490	29311401
317	R	CTTGACCTCAGTATTCTGTACTCACTCCCC	29311402	29315522
319	F	GCCCCTGAAGATATAGCACAGTCTTGCAAG	29317749	29319160
320	R	CTTGCACGTCTCCATTCATCTTGACTAAAG	29319161	29319272
321	F	TCAAACCATACAGCTGAATATGGAGAGAAG	29319273	29320195
324	R	CTTAAATCTAATAAGATGAAGGAAAATAAC	29321894	29331157
325	F	CACAGCTCCTGCACATCTGGGGAAG	29331158	29332122
326	R	CTTCCAACACCAAATGTGGTTCAATTACGC	29332123	29347103
327	F	CCATTTAAGGCCTGCCCAGCTCAAG	29347104	29352141
330	F	AGATCTGTGCAACTTTCTTCAAGCCACAAG	29356572	29358986

331	R	CTTGGATGCCTTTGATCCCCTTATAACTTA	29358987	29363174
342	F	TCACCACATGTTGAACTCTGGATGGGAAAG	29384407	29389154
345	R	CTTCTACGTAGTAGATTTTTATATGGAATT	29401642	29402954
346	F	TTTGGTTTGGTTTTCTGGGTTGAGGGGAAG	29402955	29404268
349	R	CTTCCCTCCTGGGAATGCTGTGTGG	29406181	29407971
350	F	AGCCTCCAGCTAAATGCCAACAAATGAAAG	29407972	29411033
351	R	CTTAGGATAGGAAGGAATTAGTGATACAGT	29411034	29411488
353	F	CAAATACAAAAAATCTGCCAAACAACAAAG	29411528	29412802
354	R	CTTGGAAACTGCACCATGGAACTCTCACTT	29412803	29413667
355	F	CCCTCTTCTCTTTCAAAGATGGAAAGTAAG	29413668	29419014
359	R	CTTTCTTGAGAAGGAGTTTTCATTCTAGTA	29425367	29425485
361	F	TTACTTATTTGTGACATGGTATCTAAAAAG	29426352	29436153
362	R	CTTCAGGGGAGTGAAAGAATTAAGATTTCT	29436154	29438524
366	R	CTTGCCCCCACAGGGCAGGC	29454397	29457357
367	F	GACCACATCCTGTCTAAACCCTGCCAAG	29457358	29463943
370	R	CTTTGGTATGAGGTAATAAAAATAATTGAA	29466664	29471070
371	F	CTGGCTGTCCTTTGCCACCAAACAAAG	29471071	29483618
372	R	CTTCGTGGTTGTATGCCTGTAATTGTGTTT	29483619	29486276
373	F	GTTCTTTCAGTGATAGATGAGAAGTGTAAG	29486277	29487932
375	R	CTTTCAAGGTTGCAGGTTACAAAATTATTA	29492359	29495456
376	F	TCAAGGAAGGCTACAAGAGGAGAGGCAAG	29495457	29497960
378	R	CTTTGTGAGGACGTGAGTCCGGCTG	29508039	29514012
379	F	GGATATTGTCTCCAATGGTTATGTTAAAAG	29514013	29518402
382	R	CTTTTGGATGCCATTAGCTAGACCTGAGTT	29519437	29521324
383	F	GTTGGCCTGCTATTAATCTGCTTCTACAAG	29521325	29521533
385	R	CTTAACTTACTTCACTAGATTTACTTCTGG	29527218	29527530
386	F	GTAGGCTTGAGCACTTACCAGCAGTGAAAG	29527531	29527778
387	R	CTTTCTACAAACTGGGAAAACCAGCCTTTG	29527779	29530051
388	F	CAGCTGAGAAGACCCAGCACAAATCCAAG	29530052	29530237
390	R	CTTATGAAGTTTTTCAGTTAAAAGTCACAT	29531243	29534953
391	F	AGAGCCACAGTGAACAATGTCTCTAGCAAG	29534954	29536943

393	R	CTTTTACTGGTTTAATATCCTTCCAGGCTT	29537971	29539101
394	F	CTAAAGGTCTTCTTGAGAGGAGCACTGAAG	29539102	29543893
395	R	CTTTACTTTCTGTCACCTTAAAGGTGAAAA	29543894	29545876
399	F	CAGGCTTCTGCCTGCAGAACCAAG	29555730	29559300
400	R	CTTCTCCTAGCACCAACCTTATGATCCTGG	29559301	29560638
402	F	GGAGTAAATGCTACATGACTGCCTGGGAAG	29560665	29560851
403	R	CTTGGTCTTTAATGTAGGCATGATGGGGTA	29560852	29561928
404	F	TAGTTATCCCGTGAGAAAGGTACACTCAAG	29561929	29563972
405	R	CTTGGCAGTTTGATTAATGCATGCTTAAGC	29563973	29566252
407	R	CTTGGGTACCAGTTGCATAAGCGTGG	29567141	29570404
408	F	GACCTGACACTGTTGATGACATTAACAAAG	29570405	29571012
409	R	CTTCCCAACCAAGGTGGGTGGG	29571013	29574785
410	F	CCTCACAACTTCTTTGTAGCAAGTGATAAG	29574786	29577091
411	R	CTTGCGAACAGAATAAAGGACGCATTTACC	29577092	29577561
412	F	ATTATTTTAATGTAGTTAGTAATTTGAAAG	29577562	29579620
414	R	CTTAGCTTCCATTTGTTGGGAAGAGTGGTG	29582421	29582838
415	F	TATATATACTTATGAACATGTTTGTAAAAG	29582839	29584474
417	R	CTTACTTCAGAATTAGGAAAACACAAAGCA	29584537	29585000
418	F	AGCAAAAGCATTAAGTGTAGCAGTGGAAAG	29585001	29590052
419	R	CTTTCCATTCTGAGTCTAGTGACTTAAAGG	29590053	29595530
421	F	TACCCTTATCAGAATGAAGTGTGCACAAAG	29595576	29595777
425	R	CTTGTCACTGCCCATTACCTGACTGTGC	29606038	29607875
426	F	CTATGGATGTTAACCAATAAGCTACAGAAG	29607876	29613408
429	R	CTTCTTATTGGAAAAATTGAAATTTTTCCT	29622314	29623626
431	F	ACATCCATAAGATTATTTCCGATGATCAAG	29626315	29628597
432	R	CTTCATATCTAACATTTGACTCATTGAGAG	29628598	29629851
435	F	TGCTCTACCATGCGTGAAGGTGATATAAAG	29632212	29635239
436	R	CTTAGAACATTAGGAATCATCACTAACTCT	29635240	29637005
437	F	ATAAATCTAACAAACTAAAAATCAGTCAAG	29637006	29638519
438	R	CTTGCTGCCCTCACCTGTGTTGGTAA	29638520	29640675
439	F	CTCATTTCATTTACCTCACTCCTTTAGAAG	29640676	29641133

440	R	CTTGCTTTTGTTGTAGGGATTTTACTTCCT	29641134	29642803
441	F	CATACAAAAGCCAGTTTTCTCAATGTGAAG	29642804	29649071
444	R	CTTGCCACAGCCTAGTTTGAGCCTTAGG	29652798	29652907
447	F	TTCACATGTAACAAAGCCAAATTTATGAAG	29656939	29658283
453	R	CTTGAGTGAGGCTTTATTTTTAGCATTTGG	29677935	29691941
457	F	GAGCCACTGCAGGGCTGGAAG	29701562	29703903
458	R	CTTCAATAAACAAGTGAGGGCGGGTGCA	29703904	29705290
459	F	ACATTTCTGTTGCAGGGCTATTCATGGAAG	29705291	29706210
460	R	CTTAGAATACTTAGTACTTGGTTTACACCT	29706211	29709705
461	F	GGTAACAAGAATAGCATTGAAAATTCTAAG	29709706	29713030
463	R	CTTGGCACTCTTTCAGGAAAGAAAGGACAG	29713093	29716181
464	F	CACCAGCCACCATCAACTACACAAAGAAG	29716182	29720154
465	R	CTTTGGTGTGTAGTACTTTGGAGTGATAAG	29720155	29727114
466	F	CTGAATTTAGGTGTATGTTTACTCAGTAAG	29727115	29730837
468	R	CTTTGGGGATCCTCTGTAAGTGGTTGCTC	29741658	29748329
470	F	CATAATAAAACTGAAAGGAGATGACCCAAG	29750154	29752052
472	R	CTTTCTCAATTGATTCTGAAAGGGAAAATG	29755324	29755456
474	F	TAGTACTTTAGCTAATGGCATTCACTGAAG	29757662	29764922
476	R	CTTCCTCGTTCCAGATGGGGTCTGG	29766525	29768740
477	F	TGAAATGTCCCACAAAGGATTATCAGGAAG	29768741	29769326
478	R	CTTGAGGTTTTCATATCAACAAGGCTCAGT	29769327	29770690
479	F	GGTGAGCTAGTCAGAGCAGTGCTGAAAAG	29770691	29772235
481	R	CTTGTCTTTGAGAGCCCGGTGCCT	29774284	29777533
482	F	TGGCCATTTTATATCTACTTAGGAGAAAAG	29777534	29781541
483	R	CTTTTTCCAAGTCAGGTTAGTAAAAGCAGA	29781542	29782720
484	F	GACCCCAAGGCACCCAACTTCAAG	29782721	29784642
487	F	TTCTGAGCTCCTGTTTTCCCTCAGAGTAAG	29786563	29787575
488	R	CTTATACCTTTATGACCCATAATGCACAGA	29787576	29797222
489	F	GGTTGGAGAGTTTGGAGGCTGAACACAAG	29797223	29805362
490	R	CTTGATTTCTCAAACTACTTATTGATTCGT	29805363	29806188
491	F	GTTTGTTGTGTAGGGGATGTGATCCAAAAG	29806189	29807449

492	R	CTTTCTAAAAAGGGAGGAAGGAAATCCAAG	29807450	29809879
493	F	ATTTGCCACTCAAAATCTGCACTTTCCAAG	29809880	29815593
494	R	CTTCCCTGCTGTAGGGGAGAGCG	29815594	29816516
495	F	GGAGGGGGGAGGTGGGAACAAG	29816517	29820938
496	R	CTTGATGAAGATCTTGGCATGGCAATGCAC	29820939	29835163
497	F	AAGACTCAGTTATACCAATGGTTCAATAAG	29835164	29835981
498	R	CTTCTTTGCGGAATTCCTAGGACGCTAATG	29835982	29836156
499	F	GGTGCAGGGAAAGTTGATAAGGGCAAAG	29836157	29838218
500	R	CTTGCCGCCAGGAAGCTAATTCCTC	29838219	29839237
504	R	CTTTCAGAAATGGAAGACTTTTTTAATTCT	29844732	29845798
505	F	TTCAGCTTTGTCTTGGTGTGTGGACTAAAG	29845799	29846398
506	R	CTTGAATGAGGAAACATAGGCTGAGAGGCC	29846399	29848887
507	F	GTACTGAGCCTAGCAGAGGAAGCTCAAG	29848888	29849213
508	R	CTTTCTACAGCCATGCCATTTACAGAAGCC	29849214	29850560
510	R	CTTCTACCTGCTAACAAACTTTCTCCTGTC	29851352	29854911
511	F	AGTGACCTAGATCTTCCCCTACACTTTAAG	29854912	29855302
512	R	CTTGCGCCTTTCCAGCAGTCTTGAAACATA	29855303	29859544
513	F	CATCATCTGTGCCATTTCTTAGCCCCAAAG	29859545	29863829
514	R	CTTATCCACCATGGCCCCAAGATTATCTTT	29863830	29864649
517	F	CTCTGAAACATGCAACCACACAGGCAAAG	29870040	29874795
518	R	CTTAACCAAGTAACACTACCAACTGGAGAC	29874796	29884457
521	F	TTTTTTTTCCCATTAGGCGAAATTTTAAAG	29896489	29896823
522	R	CTTCATCTGACCTCGTTTGCACGAAGCTC	29896824	29897292
523	F	TAAAACTTGCAGCAACTTTCCTGTGTTAAG	29897293	29900118
525	F	AATATGGTTTCCTTATGATGTAGCTTTAAG	29902876	29906098
526	R	CTTCCAAGTTTCTGAAGCATCCTCACCAGA	29906099	29907124
527	F	TTGTGATCAAATTTACATGTCTAAGCTAAG	29907125	29909799
529	R	CTTTTGACTAGTGGACTTTATCTGCTCTCA	29911938	29913290
530	F	TTCCTTTAGCTTTATATATTGATGGAGAAG	29913291	29913607
532	R	CTTCAGTTATACAAAGTTCTTTCATGCCCC	29913932	29918392
533	F	AGATCTCTCTACTTGTTGATTAATAGCAAG	29918393	29927218

535	R	CTTAGGTATAAAAGAAAATCTTTAAATACC	29931042	29938025
536	F	AAGTTAATTGAATGAAAATGATCAACTAAG	29938026	29947212
537	R	CTTTCTGACCAAAGTTGACAACAGCACTAT	29947213	29949609
538	F	CATGTGTTGTAGAGGCTGGGAACCACAAG	29949610	29953599
539	R	CTTTTGTTTATGGTCTGGAATGTACCCATT	29953600	29958998
540	F	CATGTGGGTTAACAGTGAGTTAAGCCCAAG	29958999	29960405
542	R	CTTCAGCCCCTTGCCTTGATGCC	29960412	29963242
543	F	TAGACATGTGAATTTTATTCTTGGAGGAAG	29963243	29963692
544	R	CTTTGAAAGACCCATGTCATAGTACGTGTT	29963693	29963925
546	F	GATTTTTTTTTCCTAGGGGTTTATTTCAAG	29967853	29970155
547	R	CTTAGGCTCAAAGATGACTGCAGAGGAGAG	29970156	29977864
548	F	AAGGAAAAAAAAAAGAATAAGGTTTCAGAAG	29977865	29978715
549	R	CTTGCAAGCTAGCTAGCCCAAGGGATAC	29978716	29979319
550	F	TGTACAGTATAAGCACCCCTATAGCCAAAG	29979320	29988808
551	R	CTTTCAATTTTTCCACATCCTGCCCAACAC	29988809	29991479
553	R	CTTTCCAGTATTGCTGGAGCTGTAGTCCGA	29995277	29996459
554	F	GCAACTCAGTGAAATTAACCAAAGATGAAG	29996460	30001398
555	R	CTTTCCAGGCATAAAGCAGAGACAGGCAAA	30001399	30005000

***USP22* (chr.11; mm9)**

1	R	CTTTGAGACTAGACCGAAGTCTCCAGAATC	60917307	60918190
2	F	GGGAGGAAGATAAAAAGATGGGGATGGAAG	60918191	60924238
5	F	CTGTCAGTCTCCTGCTGCCACTAACAAG	60932406	60937638
6	R	CTTGAGGAGACAGCACTGCTGGTAGATAG	60937639	60942991
9	F	GTGTCAAGGAGGCAGACCTTCAGGAAG	60949684	60966226
10	R	CTTACTGTGGGCTACCCATTTGTACTCTTA	60966227	60970541
11	F	AACCCATCTCTAACAATCCCTTTGTAAAAG	60970542	60976197
12	R	CTTTTATTTATAAGAGATCTTAGCTAATGA	60976198	60977387
14	F	GCAGGAGTCTAAGCCACCAGGGAAG	60977797	60977935

16	R	CTTTTTTCCAAATAGTAACCCACAGGGCCA	60981479	61000571
17	F	ACTTTTCCCAAAATCCAAGCTGACTTCAAG	61000572	61003268

Table S 2 Mouse 5C primers for Shh and USP22 regions

Publication

Regulatory Domains and Their Mechanisms

Nezha S. Benabdallah[1,2] and Wendy A. Bickmore[1]

[1]*MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH42XU, United Kingdom*

[2]*Edinburgh Super Resolution Imaging Consortium, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH42XU, United Kingdom*

Correspondence: wendy.bickmore@igmm.ed.ac.uk

The concept of gene regulation is being refined as our understanding of the role of enhancer elements grows. Although described more than 30 years ago, the mechanisms through which these *cis*-regulating elements operate remain under debate. With the recognition that most of the human genetic variation contributing to common disease risk lies outside of genes and probably in enhancers, unraveling these mechanisms becomes ever more important. Originally, a popular view was to consider regulatory elements as an entry site for the transcription machinery that could scan the intervening chromatin until the cognate core promoter was located. Now, the most prominent model for distal enhancer–promoter interaction involves direct enhancer/promoter contacts with a looping out of intervening chromatin. However, a rising awareness of the importance of chromatin architecture and organization forces us to consider enhancer–promoter communication in light of the polymer folding properties of chromatin. Here, we discuss how three-dimensional chromatin folding, topological domains, and the constrained motion, plasticity, and accessibility of chromatin could offer a structural basis for regulatory domains that greatly enhances the probability of enhancer–promoter and transcription factor–promoter interactions and gene activation.

The definition of a gene has evolved significantly over the past 35 years. In metazoans, and particularly in mammals, the functional genetic unit now goes far beyond the genomic region that encodes an mRNA. 5′ elements adjacent to the transcription start site encompass the core promoter and serve as the docking site for the RNA polymerase II preinitiation complex (PIC). However, additional layers of complexity have emerged, including multiple alternative promoters, alternative exons, and *cis*-regulatory sequences positioned upstream, downstream, or even within an intron of the transcription unit.

Although the core promoter determines where transcription begins, the *cis*-regulatory sequences—or enhancers—stimulate promoters in a time and tissue-specific manner and increase the efficiency of transcription in a position- and orientation-independent manner. However, both enhancer and gene activation are associated with similar events—that is, recruitment of transcription factors, coactivators, and RNA polymerase II and production of RNA (albeit short and unstable at enhancers) (Koch et al. 2011). Therefore, enhancers have long been seen as simple extensions of promoters as both lead to assembly of an active transcription complex and are often bound by the same components (Andersson et al. 2015).

The first enhancer described was the 72-bp tandem repeat of SV40 DNA, which was identified as a sequence located ~100 nt upstream of the core early viral promoter (Fig. 1) and whose deletion reduced gene expression and abolished virus viability (Benoist and Chambon 1981; Gruss et al. 1981). Later, the SV40 enhancer was found to comprise three main segments, each of which provides cell type specificity (Schirm et al. 1987). In parallel, it was discovered that the SV40 enhancer could work when associated with other diverse promoters. Surprisingly, it could also act over considerable distances (>10 kb) and its activity appeared to rely on the sequence of the intervening chromatin rather than the distance (Banerji et al. 1981; Mellon et al. 1981; Moreau et al. 1981).

Other sequences with equivalent function in mammalian cells were soon identified (Conrad and Botchan 1982; Banerji et al. 1983), including locus control regions (LCRs)—which define regulatory domains that are groups of enhancers. The first LCR element identified 25-kb upstream of the human ε-, γ-, δ-, and β-globin genes (Fig. 1) shared characteristics of the SV40 enhancer including clustered DNase I hypersensitive sites (DHSs) (Grosveld et al. 1987; for review, see Li et al. 2002).

Now, human and mouse Mendelian genetics have revealed enhancers located up to 1 million base pairs (megabase/Mb) 5′ or 3′ of target genes (Fig. 1), often with other intervening genes located in-between, or embedded within the introns of nearby genes (Noonan and McCallion 2010). Thus, the functional unit of gene regulation in the mammalian genome, for coding regions that are themselves often <10 kb in size, can be of the order of 2 Mb in size.

Since they were first described, many of the molecular characteristics of enhancers have been established including transcription factor binding, coactivator recruitment, DHS, distinctive histone marks, and eRNA production. Here, we specifically consider the regulatory domains in which enhancers operate and how these domains may be

 doi: 10.1101/sqb.2015.80.027268

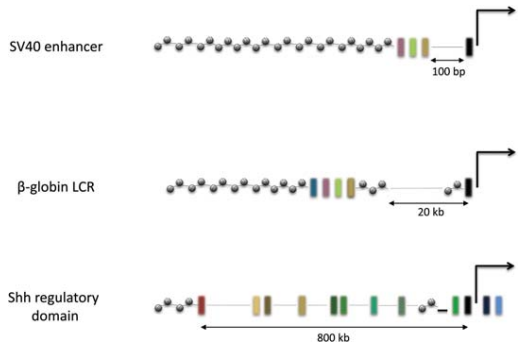

Figure 1. Different scales of enhancer and promoter distances. (*Top*) Proximal enhancers can be located a few tens of bp from their target promoter. An example of this is the SV40 early promoter. (*Middle*) Enhancers such as the β-globin locus control region (LCR) are located tens of kilobases from their target promoters. (*Bottom*) Some developmental enhancers can be located hundreds or even 1000 kb from their target gene, often with other, unrelated genes, in between. The *Shh* limb enhancer (ZRS) is an example of this.

integrated in specific chromatin conformations that influence enhancer–promoter communication.

cis-ACTING ELEMENTS

For the most part, enhancers function in *cis*. Exceptions include the ability of enhancers to work between homologous chromosomes in *Drosophila* cells that show somatic pairing (Schoborg et al. 2013) and the infrequent interaction of multiple enhancers—located on different chromosomes—with murine olfactory receptor genes, as the basis for stochastic choice in olfactory receptor gene expression in olfactory neurons (Markenscoff-Papadimitriou et al. 2014).

The requirement for a target gene and its enhancer to be located in *cis*, in order for gene activation to be robust, is evidenced by the Mendelian genetic disorders that arise when translocations separate enhancer and promoter onto different chromosomes (Kleinjan and van Heyningen 2005; Benko et al. 2009; Lango Allen et al. 2014; Rainger et al. 2014).

Genetic evidence indicates that enhancers can act over 1 Mb of intervening chromatin. However, their range may not be much greater than this. Chromosomal inversions that increase the genomic distance between an enhancer and its target gene can, in some cases, abrogate enhancer function (Lettice et al. 2011).

Chromosome structural rearrangements have set the likely limits of enhancer function in the human genome and indicate that correct enhancer function usually depends on the enhancer being present within a range of ~1 Mb in *cis* with its target gene. However, it is important to remember that enhancers are not constrained to have to work at a distance. Both enhancer reporter assays (Bhatia et al. 2015) and localized transposon hopping (Anderson et al. 2014; Symmons et al. 2014) show that endogenous mammalian enhancers that act at long range in their normal chromosomal context are quite capable of also acting at short range. It is within this framework that we now consider possible mechanisms of enhancer function in the context of three-dimensional chromatin conformation.

SPREADING OF A SIGNAL: LINKING AND TRACKING MODELS

The original view of regulatory element action was that it provided a precise docking site for RNA polymerase II or other components of the transcriptional machinery, followed by tracking of these factors on the chromatin fiber until they met the cognate core promoter and activated it in a time- and position-dependent manner.

Exploration of tracking-like mechanisms at several loci, such as β-globin, has reported the unidirectional spreading of histone H3 and H4 acetylation, the CBP/p300 histone acetyltransferases, P/CAF, subunits of the SWI/SNF chromatin remodeling complex, or RNA polymerase II and TBP with accompanying synthesis of short, polyadenylated, intergenic RNAs (Gribnau et al. 2000; Hatzis and Talianidis 2002; Spicuglia et al. 2002; Masternak et al. 2003; Kim and Dean 2004; Zhao and Dean 2004; Wang et al. 2005; Zhu et al. 2007). These studies implicate tracking as the primary step for enhancer–promoter interaction and speculate that a stable enhancer–promoter loop is only formed when the tracking is complete (Fig. 2A). However, the formation of stable enhancer–promoter complexes has not always been convincingly shown. The linking hypothesis differs from the tracking mechanism, as the proteins involved do not scan all of the intervening chromatin but instead oligomerize to spread the signal (Fig. 2B).

In these examples of enhancer–promoter communication, the genomic distance separating the two elements is modest (generally <20 kb) and, from both kinetic and spatial perspectives, it is hard to imagine how one-dimensional tracking or linking mechanisms could operate over much larger (e.g., 1-Mb) distances and over intervening genes.

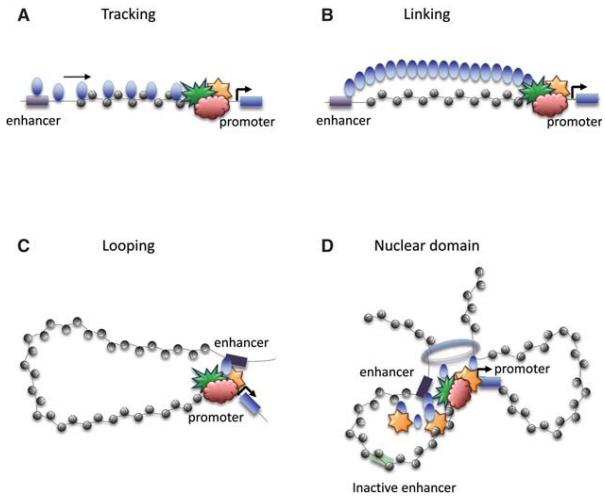

Figure 2. Enhancer and promoter communication mechanisms. Activating signals, such as transcription factors, bind enhancers recruiting other proteins, such as coactivators, that may track along the chromatin fiber until they meet their cognate promoter (*A*: tracking model) or oligomerize and form a protein bridge to meet their cognate promoter (*B*: linking model). Enhancer elements may be brought to the vicinity of their cognate promoter by direct interaction and looping out of the intervening chromatin region (*C*: looping model), or enhancers may just be constrained within the same nuclear domain as their target genes with protein and chromatin constrained diffusion then being sufficient to activate gene expression (*D*: nuclear domain model).

BRIDGING THE DIVIDE: THE CHROMOSOME LOOPING MODEL

The current favored model for long-range enhancer–promoter communication involves direct homotypic or heterotypic interactions between enhancer- and promoter-bound proteins to form a chromatin loop that juxtaposes enhancer and promoter at the loop base and that loops out the intervening chromatin (Fig. 2C; Su et al. 1991).

Elegant chromosome engineering experiments at the β-globin locus have shown that forcing a chromosome loop between an enhancer and its target promoter is sufficient to trigger gene expression in erythroid cells, in which some other key transcription factors required for β-globin expression are already bound (Deng et al. 2012, Deng et al. 2014). However, the extent to which endogenous enhancers and promoters form relatively stable chromatin loops is currently unclear.

Much of the evidence for enhancer–promoter loops comes from the detection of enhancer–promoter ligation products in the chromosome conformation capture (3C) technique and its derivatives (4C, 5C, Hi-C, etc.) (Tolhuis et al. 2002; Palstra et al. 2003; Montavon et al. 2011; Jin et al. 2013). However, such assays do not have the ability to determine the frequency at which such interactions take place, to determine the molecular nature of such interactions, or to determine whether the interactions are the cause or consequence of enhancer-driven gene activation.

In some instances fluorescence in situ hybridization (FISH) has shown the spatial juxtaposition of distant enhancers and promoters at frequencies high enough (30% or more of alleles) to be interpreted as the base of a chromosome loop (Williamson et al. 2012; Lettice et al. 2014). However in other cases, visual assays do not provide evidence for relatively stable chromosome loops (Williamson et al. 2012). This does not exclude the existence of very transient enhancer–promoter contacts, but we would argue that such infrequent contacts should not be described as chromatin loops.

How enhancers and promoters find each other in nuclear space is not known, but it seems most likely that it occurs as a consequence of the rapid, but constrained, diffusion (diffusion coefficient of $\sim 1 \times 10^{-4}$ $\mu m^2/sec$) that characterizes chromatin motion in living human cells (Chubb et al. 2002). This implies that enhancers and their cognate promoters must both be located within a small enough nuclear volume to be compatible with the radius of constraint for chromatin motion. For the genomic loci measured to date, this constraint (<0.5 μm) could encompass loci that are up to ~1 Mb from each other (Fig. 3A,B).

The formation of a stable chromatin loop has an entropic cost to polymer dynamics as it would need to overcome inherent diffusion. It would be interesting to investigate the strength of binding energies between enhancer- and promoter-bound proteins that would be required to stabilize chromatin loops of various sizes. This may be even more of an issue for enhancers located close to promoters, where the persistence length of the chroma-

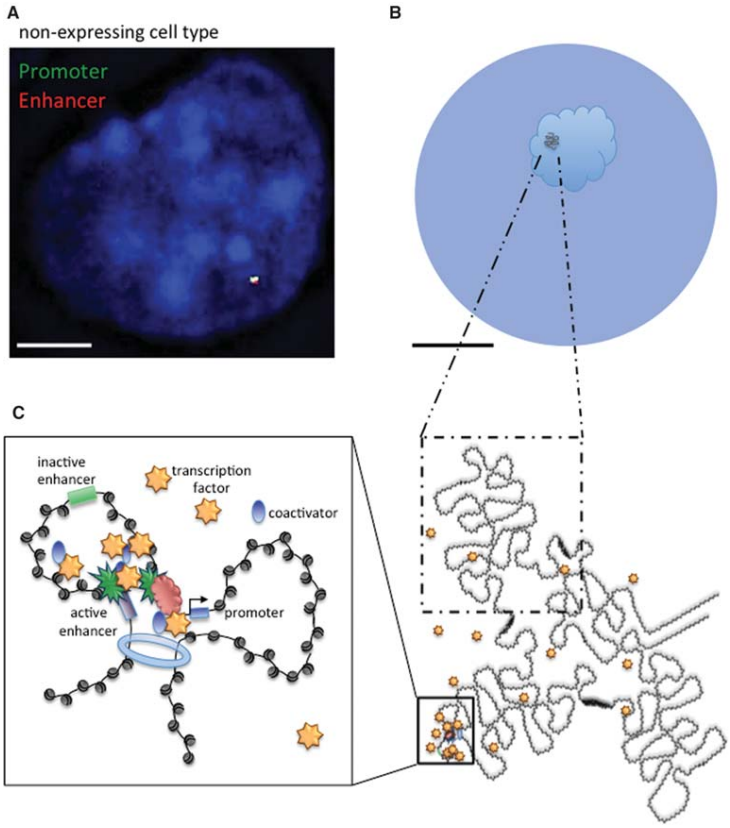

Figure 3. Chromatin compaction and constrained motion and its influence on transcription factor diffusion. (*A*) FISH image illustrating a highly compact chromatin conformation of an inactive enhancer (red) and its cognate promoter (green), despite their being separated by several hundred kilobases, in the cell nucleus of a nonexpressing cell type. Scale bar, 5 μm. (*B*) Schematic representation of constrained chromatin motion (0.5 μm average range) where chromatin can only sample a small proportion of the nuclear volume or indeed of the chromosome territory (pale blue). Enhancers and their cognate promoters must be close enough to each other in nuclear space to be able to encounter each other through constrained diffusion. Scale bar, 5 μm. (Chubb and Bickmore 2003.) (*C*) Compact chromatin domains, such as topologically associated domains (TADs), and their viscoelastic properties, can affect the dimensionality of transcription factor diffusion and create compact volumes that can locally concentrate transcription factors, RNA polymerase, and coactivators.

tin polymer may be a further energetic barrier to loop formation.

KEEPING IT CONCENTRATED: REGULATION WITHIN NUCLEAR DOMAINS

The activation of gene expression requires the formation of protein–protein and protein–DNA complexes, the efficiency of which will be dependent on the laws of mass action and so will be very sensitive to local concentration of both proteins and DNA. Chromatin packaging may therefore be an important determinant of regulatory domains. In vivo analysis of the binding regimes of transcription factors shows that some use facilitated diffusion (i.e., 3D diffusion interspersed with local 1D sliding along the chromatin fiber). This greatly reduces their search space to give an efficient means of finding their specific binding sites (Chen et al. 2014; Izeddin et al. 2014; Woringer et al. 2014; Normanno et al. 2015). A transcription factor bound at an enhancer might then rather efficiently, by oversampling the local volume, be able to find a promoter-proximal binding site located

within the same chromatin domain, obviating any need for a specific enhancer–promoter chromatin loop (Figs. 2D, 3C).

A second mode by which chromatin packing might affect enhancer function is by reducing the volume in which proteins involved in transcription (e.g., coactivators, components of the PIC) are free to diffuse and to form complexes with each other (excluded volume effect) (Woringer et al. 2014). Interestingly, ~1 Mb chromatin domains containing genes subject to complex developmental regulation that is driven by suites of enhancers appear to have a very compact chromatin structure, even in nonexpressing tissues (Fig. 3A; Williamson et al 2012; Lettice et al. 2014). This suggests that they may have evolved to facilitate both excluded volume and facilitated diffusion effects. Because of the viscoelastic properties of the nuclear environment, compact chromatin domains have been shown to affect chromatin motion and to dramatically reduce the first encounter time for distant genomic regions (Lucas et al. 2014).

A domain model for enhancer function may help to explain how primary and shadow enhancers, by both being contained in the same compact nuclear domain, could provide robustness to gene regulation during development, particularly when key molecules required for specific gene activation are limiting in the nucleus (Bothma et al. 2015; Lam et al. 2015). It also provides a plastic environment within which new enhancers can evolve to drive new patterns of gene regulation, or within which existing enhancers can relocate (Villar et al. 2015), without having to disrupt and rewire preexisting defined chromatin loops mediated by other enhancers in the same domain.

How such chromatin domains remain compact is unclear. It has been suggested that closed DNA domains could be formed by supercoiling with architectural proteins bound at the ends of these domains hindering the axial rotation that would relieve supercoiling (Duncan et al. 1994; Ner et al. 1994; Yang et al. 1995; Gilbert and Allan 2014). Indeed several studies have shown that enhancers and their target promoters tend to be located within the same topologically associated domains (TADs)—as defined from HiC data (Jin et al. 2013). In addition, reporter assays show that the influence of an enhancer declines sharply beyond the TAD domain boundary (Anderson et al. 2014; Symmons et al. 2014). Conversely, disrupting TAD boundaries can rewire enhancer and promoter communication leading to inappropriate gene activation (Lupiáñez et al. 2015).

What forms TAD boundaries is not completely clear, but convergently oriented CTCF sites, which also recruit cohesin, have been shown to be one important factor (Guo et al. 2015). The mechanistic basis for this is unclear, but it suggests the propagation of a directional signal from the boundary CTCF site in toward the TAD. Conceptually, this has some similarity to linking/tracking models of enhancer–promoter communication. However, the nature of that signal is unknown and it is likely that not all boundaries are the same.

CONCLUSION

Characterization of mammalian genomes using chromosome conformational capture techniques and microscopic imaging strongly support a model in which three-dimensional chromatin structures are organized into domains within which DNA–DNA and protein–DNA interactions are facilitated. It is within this framework that possible mechanisms of enhancer activation must be considered. Therefore, it becomes imperative that methods are developed that will allow the biophysical properties of chromatin domains and of transcription factor and chromatin motion to be quantified and manipulated in physiologically relevant systems.

ACKNOWLEDGMENTS

N.S.B. is supported by a Ph.D. studentship from the Edinburgh Super Resolution Imaging consortium (ESRIC). W.A.B. is supported by a Unit programme grant from the U.K. Medical Research Council. We thank Nick Gilbert (MRC HGU, Edinburgh) for critical reading of this manuscript.

REFERENCES

Anderson E, Devenney PS, Hill RE, Lettice LA. 2014. Mapping the *Shh* long-range regulatory domain. *Development* **141:** 3934–3943.

Andersson R, Sandelin A, Danko CG. 2015. A unified architecture of transcriptional regulatory elements. *Trends Genet* **31:** 426–433.

Banerji J, Rusconi S, Schaffner W. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27:** 299–308.

Banerji J, Olson L, Schaffner W. 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33:** 729–740.

Benko S, Fantes JA, Amiel J, Kleinjan DJ, Thomas S, Ramsay J, Jamshidi N, Essafi A, Heaney S, Gordon CT, et al. 2009. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat Genet* **41:** 359–364.

Benoist C, Chambon P. 1981. In vivo sequence requirements of the SV40 early promotor region. *Nature* **290:** 304–310.

Bhatia S, Gordon CT, Foster RG, Melin L, Abadie V, Baujat G, Vazquez MP, Amiel J, Lyonnet S, van Heyningen V, et al. 2015. Functional assessment of disease-associated regulatory variants in vivo using a versatile dual colour transgenesis strategy in zebrafish. *PLoS Genet* **11:** e1005193.

Bothma JP, Garcia HG, Ng S, Perry MW, Gregor T, Levine M. 2015. Enhancer additivity and non-additivity are determined by enhancer strength in the *Drosophila* embryo. *Elife* **4**. doi: 10.7554/eLife.07956.

Chen J, Zhang Z, Li L, Chen BC, Revyakin A, Hajj B, Legant W, Dahan M, Lionnet T, Betzig E, et al. 2014. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156:** 1274–1285.

Chubb JR, Bickmore WA. 2003. Considering nuclear compartmentalization in the light of nuclear dynamics. *Cell* **112:** 403–406.

Chubb JR, Boyle S, Perry P, Bickmore WA. 2002. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr Biol* **12:** 439–445.

Conrad SE, Botchan MR. 1982. Isolation and characterization of human DNA fragments with nucleotide sequence homologies

with the simian virus 40 regulatory region. *Mol Cell Biol* **2:** 949–965.
Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel GA. 2012. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149:** 1233–1244.
Deng W, Rupon JW, Krivega I, Breda L, Motta I, Jahn KS, Reik A, Gregory PD, Rivella S, Dean A, et al. 2014. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* **158:** 849–860.
Duncan R, Bazar L, Michelotti G, Tomonaga T, Krutzsch H, Avigan M, Levens D. 1994. A sequence-specific, single-strand binding protein activates the far upstream element of c-*myc* and defines a new DNA-binding motif. *Genes Dev* **8:** 465–480.
Gilbert N, Allan J. 2014. Supercoiling in DNA and chromatin. *Curr Opin Genet Dev* **25:** 15–21.
Gribnau J, Diderich K, Pruzina S, Calzolari R, Fraser P. 2000. Intergenic transcription and developmental remodeling of chromatin subdomains in the human β-globin locus. *Mol Cell* **5:** 377–386.
Grosveld F, van Assendelft GB, Greaves DR, Kollias G. 1987. Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* **51:** 975–985.
Gruss P, Dhar R, Khoury G. 1981. Simian virus 40 tandem repeated sequences as an element of the early promoter. *Proc Natl Acad Sci* **78:** 943–947.
Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, et al. 2015. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162:** 900–910.
Hatzis P, Talianidis I. 2002. Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Mol Cell* **10:** 1467–1477.
Izeddin I, Récamier V, Bosanac L, Cissé II, Boudarene L, Dugast-Darzacq C, Proux F, Bénichou O, Voituriez R, Bensaude O, et al. 2014. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *Elife* **3**. doi: 10.7554/eLife.02230.
Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B. 2013. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503:** 290–294.
Kim A, Dean A. 2004. Developmental stage differences in chromatin subdomains of the β-globin locus. *Proc Natl Acad Sci* **101:** 7028–7033.
Kleinjan DA, van Heyningen V. 2005. Long-range control of gene expression: Emerging mechanisms and disruption in disease. *Am J Hum Genet* **76:** 8–32.
Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, Spicuglia S, de la Chapelle AL, Heidemann M, Hintermair C, et al. 2011. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol* **18:** 956–963.
Lam DD, de Souza FS, Nasif S, Yamashita M, López-Leal R, Otero-Corchon V, Meece K, Sampath H, Mercer AJ, Wardlaw SL, et al. 2015. Partially redundant enhancers cooperatively maintain mammalian *Pomc* expression above a critical functional threshold. *PLoS Genet* **11:** e1004935.
Lango Allen H, Caswell R, Xie W, Xu X, Wragg C, Turnpenny PD, Turner CL, Weedon MN, Ellard S. 2014. Next generation sequencing of chromosomal rearrangements in patients with split-hand/split-foot malformation provides evidence for *DYNC1I1* exonic enhancers of *DLX5/6* expression in humans. *J Med Genet* **51:** 264–267.
Lettice LA, Daniels S, Sweeney E, Venkataraman S, Devenney PS, Gautier P, Morrison H, Fantes J, Hill RE, FitzPatrick DR. 2011. Enhancer-adoption as a mechanism of human developmental disease. *Hum Mutat* **32:** 1492–1499.
Lettice LA, Williamson I, Devenney PS, Kilanowski F, Dorin J, Hill RE. 2014. Development of five digits is controlled by a bipartite long-range *cis*-regulator. *Development* **141:** 1715–1725.
Li Q, Peterson KR, Fang X, Stamatoyannopoulos G. 2002. Locus control regions. *Blood* **100:** 3077–3086.
Lucas JS, Zhang Y, Dudko OK, Murre C. 2014. 3D trajectories adopted by coding and regulatory DNA elements: First-passage times for genomic interactions. *Cell* **158:** 339–352.
Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161:** 1012–1025.
Markenscoff-Papadimitriou E, Allen WE, Colquitt BM, Goh T, Murphy KK, Monahan K, Mosley CP, Ahituv N, Lomvardas S. 2014. Enhancer interaction networks as a means for singular olfactory receptor expression. *Cell* **159:** 543–557.
Masternak K, Peyraud N, Krawczyk M, Barras E, Reith W. 2003. Chromatin remodeling and extragenic transcription at the MHC class II locus control region. *Nat Immunol* **4:** 132–137.
Mellon P, Parker V, Gluzman Y, Maniatis T. 1981. Identification of DNA sequences required for transcription of the human alpha 1-globin gene in a new SV40 host-vector system. *Cell* **27:** 279–288.
Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F, Duboule D. 2011. A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147:** 1132–1145.
Moreau P, Hen R, Wasylyk B, Everett R, Gaub MP, Chambon P. 1981. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* **9:** 6047–6068.
Ner SS, Travers AA, Churchill ME. 1994. Harnessing the writhe: A role for DNA chaperones in nucleoprotein-complex formation. *Trends Biochem Sci* **19:** 185–187.
Noonan JP, McCallion AS. 2010. Genomics of long-range regulatory elements. *Annu Rev Genomics Hum Genet* **11:** 1–23.
Normanno D, Boudarène L, Dugast-Darzacq C, Chen J, Richter C, Proux F, Bénichou O, Voituriez R, Darzacq X, Dahan M. 2015. Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat Commun* **6:** 7357.
Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. 2003. The β-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* **35:** 190–194.
Rainger JK, Bhatia S, Bengani H, Gautier P, Rainger J, Pearson M, Ansari M, Crow J, Mehendale F, Palinkasova B, et al. 2014. Disruption of SATB2 or its long-range *cis*-regulation by SOX9 causes a syndromic form of Pierre Robin sequence. *Hum Mol Genet* **23:** 2569–2579.
Schirm S, Jiricny J, Schaffner W. 1987. The SV40 enhancer can be dissected into multiple segments, each with a different cell type specificity. *Genes Dev* **1:** 65–74.
Schoborg T, Kuruganti S, Rickels R, Labrador M. 2013. The Drosophila *gypsy* insulator supports transvection in the presence of the *vestigial* enhancer. *PLoS One* **8:** e81331.
Spicuglia S, Kumar S, Yeh JH, Vachez E, Chasson L, Gorbatch S, Cautres J, Ferrier P. 2002. Promoter activation by enhancer-dependent and -independent loading of activator and coactivator complexes. *Mol Cell* **10:** 1479–1487.
Su W, Jackson S, Tjian R, Echols H. 1991. DNA looping between sites for transcriptional activation: Self-association of DNA-bound Sp1. *Genes Dev* **5:** 820–826.
Symmons O, Uslu VV, Tsujimura T, Ruf S, Nassari S, Schwarzer W, Ettwiller L, Spitz F. 2014. Functional and topological characteristics of mammalian regulatory domains. *Genome Res* **24:** 390–400.
Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. 2002. Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol Cell* **10:** 1453–1465.
Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, et al. 2015. Enhancer evolution across 20 mammalian species. *Cell* **160:** 554–566.

Wang Q, Carroll JS, Brown M. 2005. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol Cell* **19:** 631–642.

Williamson I, Eskeland R, Lettice LA, Hill AE, Boyle S, Grimes GR, Hill RE, Bickmore WA. 2012. Anterior–posterior differences in HoxD chromatin topology in limb development. *Development* **139:** 3157–3167.

Woringer M, Darzacq X, Izeddin I. 2014. Geometry of the nucleus: A perspective on gene expression regulation. *Curr Opin Chem Biol* **20:** 112–119.

Yang Y, Westcott TP, Pedersen SC, Tobias I, Olson WK. 1995. Effects of localized bending on DNA supercoiling. *Trends Biochem Sci* **20:** 313–319.

Zhao H, Dean A. 2004. An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. *Nucleic Acids Res* **32:** 4903–4919.

Zhu X, Ling J, Zhang L, Pi W, Wu M, Tuan D. 2007. A facilitated tracking and transcription mechanism of long-range enhancer function. *Nucleic Acids Res* **35:** 5532–5544.

Abbreviations

A adenine

ac acetylation

bp base pair(s)

BSA bovine serum albumin

BP Belmont peptide DELQPASIDP

C cytosine

cDNA complementary DNA

ChIP chromatin immunoprecipitation

ChIP-chip – ChIP on chip (ChIP combined with microarrays)

chr- chromosome

CO2 Carbon Dioxide

Cy3 Cyanine 3

Cy5 Cyanine 5

C-terminus Carboxyl-terminus

d nuclear interprobe distance (μm)

d^2 mean squared nuclear interprobe distance (μm^2)

DAPI 4',6-diamidino-2-phenylindole

DMSO- Dimethyl Sulfoxide

dH2O deionised water

DNA deoxyribo nucleic acid

dNTP deoxynucleotide triphosphate

DTT dithiothreitol

E embryonic stage

EDTA ethylene diamine tetraacetic acid

EGTA ethylene glycol tetraacetic acid

ENCODE Encyclopedia of DNA Elements

ESC embryonic stem cell(s)

FITC Fluorescein isothiocyanate

FCS foetal calf serum

FFT Fast Fourier Transformation (computational algorithm for performing DFTs)

FISH fluorescence in situ hybridization

G guanine

GEO Gene Expression Omnibus

GFP green fluorescent protein

H3K4 Histone H3 Lysine 4

H3K27 Histone H3 Lysine 27

HAT Histone Acetyltransferase

HDAC Histone Deacetylase

HS Hypersensitive Site

LB Luria Bertani broth

LDB1 lim domain binding 1

LCR locus control region

kb kilo base pair(s)

M molar

Mb mega base pair(s)

me methylation

MMU *Mus musculus* chromosome

MNase Micrococcal Nuclease

mRNA messenger RNA

µg microgram

µl microlitre

µm micrometre

nm nanometre

N-Terminus Amino Terminus

NaCl Sodium Chloride

NaOH Sodium Hydroxide

NPC neural progenitor cell(s)

PARP1 Poly(ADP-ribose) Polymerase 1

PBS phosphate buffered saline

PCR polymerase chain reaction

PFA paraformaldehyde

qRT-PCR quantitative reverse transcription polymerase chain reaction

RNA ribonucleic acid

RNAPII RNA polymerase II

RT Room Temperature

SBE *Shh* brain enhancer

SDS sodium dodecyl sulphate

SEM standard error of the mean

SFPE *Shh* floor plate enhancer

SNP single nucleotide polymorphism

SSC Saline Sodium Citrate

T thymine

TAD topologically associated domain

TAE Tris-EDTA acetic acid buffer

TALE Transcription activator-like effector

TE Tris-EDTA

TF transcription factor

TFBS transcription factor binding site

TSS transcription start site

UV ultra-violet radiation

WT wild-type

ZRS zone of polarizing activity regulatory sequence

3C chromatin conformation capture

5C chromatin conformation capture carbon copy

WGA Whole Genome Amplification, Kit from Sigma

References

Abranches, E., Silva, M., Pradier, L., Schulz, H., Hummel, O., Henrique, D., and Bekman, E. (2009). Neural differentiation of embryonic stem cells in vitro: a road map to neurogenesis in the embryo. PLoS One *4*, e6286.

Allan, J., Hartman, P.G., Crane-Robinson, C., and Aviles, F.X. (1980). The structure of histone H1 and its location in chromatin. Nature *288*, 675–679.

Allan, J., Cowling, G.J., Harborne, N., Cattini, P., Craigie, R., and Gould, H. (1981). Regulation of the higher-order structure of chromatin by histones H1 and H5. J. Cell Biol. *90*, 279–288.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat. Rev. Mol. Cell Biol. *16*, 155–166.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. Dev. Cell *16*, 47–57.

Anderson, E., and Hill, R.E. (2014). Long range regulation of the sonic hedgehog gene. Curr. Opin. Genet. Dev. *27*, 54–59.

Anderson, E., Devenney, P.S., Hill, R.E., and Lettice, L.A. (2014). Mapping the Shh long-range regulatory domain. Development *141*, 3934–3943.

Augui, S., Nora, E.P., and Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. Nat. Rev. Genet. *12*, 429–442.

Balaskas, N., Ribeiro, A., Panovska, J., Dessaud, E., Sasai, N., Page, K.M., Briscoe, J., and Ribes, V. (2012). Gene Regulatory Logic for Reading the Sonic Hedgehog Signaling

Gradient in the Vertebrate Neural Tube. Cell *148*, 273–284.

Bancaud, A., Huet, S., Daigle, N., Mozziconacci, J., Beaudouin, J., and Ellenberg, J. (2009). Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. EMBO J. *28*, 3785–3798.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308.

Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell *33*, 729–740.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in Drosophila. Cell *144*, 214–226.

Bartman, C.R., Hsu, S.C., Hsiung, C.C.-S., Raj, A., and Blobel, G.A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. Mol. Cell *62*, 237–247.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387–396.

Belloni, E., Muenke, M., Roessler, E., Traverse, G., Siegel-Bartelt, J., Frumkin, A., Mitchell, H.F., Donis-Keller, H., Helms, C., Hing, A.V., et al. (1996). Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. Nat. Genet. *14*, 353–356.

Belmont, A.S., Dietzel, S., Nye, A.C., Strukov, Y.G., and Tumbar, T. (1999). Large-scale chromatin structure and function. Curr. Opin. Cell Biol. *11*, 307–311.

Benabdallah, N.S., and Bickmore, W.A. (2015). Regulatory Domains and Their Mechanisms. Cold Spring Harb Symp Quant Biol *80*, 45–51.

Benko, S., Fantes, J.A., Amiel, J., Kleinjan, D.-J., Thomas, S., Ramsay, J., Jamshidi, N., Essafi, A., Heaney, S., Gordon, C.T., et al. (2009). Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. Nat. Genet. *41*, 359–364.

Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promoter region. Nature *290*, 304–310.

Berlivet, S., Paquette, D., Dumouchel, A., Langlais, D., Dostie, J., and Kmita, M. (2013). Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. PLoS Genet. *9*, e1004018.

Bhatia, S., Bengani, H., Fish, M., Brown, A., Divizia, M.T., de Marco, R., Damante, G., Grainger, R., van Heyningen, V., and Kleinjan, D.A. (2013). Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. Am. J. Hum. Genet. *93*, 1126–1134.

Bhatia, S., Gordon, C.T., Foster, R.G., Melin, L., Abadie, V., Baujat, G., Vazquez, M.-P., Amiel, J., Lyonnet, S., van Heyningen, V., et al. (2015). Functional Assessment of Disease-Associated Regulatory Variants In Vivo Using a Versatile Dual Colour Transgenesis Strategy in Zebrafish. PLOS Genet. *11*, e1005193.

Bickmore, W.A., and van Steensel, B. (2013). Genome Architecture: Domain Organization of Interphase Chromosomes. Cell *152*, 1270–1284.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E.M. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat. Genet. *44*, 148–156.

Bothma, J.P., Garcia, H.G., Ng, S., Perry, M.W., Gregor, T., and Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. Elife *4*.

Boyle, S., Rodesch, M.J., Halvensleben, H.A., Jeddeloh, J.A., and Bickmore, W.A. (2011). Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. Chromosome Res. *19*, 901–909.

Brown, J.M., Leach, J., Reittie, J.E., Atzberger, A., Lee-Prudhoe, J., Wood, W.G., Higgs, D.R., Iborra, F.J., and Buckle, V.J. (2006). Coregulated human globin genes are frequently in spatial proximity when active. J. Cell Biol. *172*, 177–187.

Brown, J.M., Green, J., das Neves, R.P., Wallace, H.A.C., Smith, A.J.H., Hughes, J., Gray, N., Taylor, S., Wood, W.G., Higgs, D.R., et al. (2008). Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. J. Cell Biol. *182*, 1083–1097.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Bulger, M., and Groudine, M. (1999). Looping versus linking: toward a model for long-distance gene activation. Genes Dev. *13*, 2465–2477.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327–339.

Bundred, N., Gardovskis, J., Jaskiewicz, J., Eglitis, J., Paramonov, V., McCormack, P., Swaisland, H., Cavallin, M., Parry, T., Carmichael, J., et al. (2013). Evaluation of the pharmacodynamics and pharmacokinetics of the PARP inhibitor olaparib: a phase I multicentre trial in patients scheduled for elective breast cancer surgery. Invest. New Drugs *31*, 949–958.

Cai, H., and Levine, M. (1995). Modulation of enhancer-promoter interactions by insulators in the Drosophila embryo. Nature *376*, 533–536.

Cai, C., Thorne, J., and Grabel, L. (2008). Hedgehog serves as a mitogen and survival factor during embryonic stem cell neurogenesis. Stem Cells *26*, 1097–1108.

Cairns, B.R. (2009). The logic of chromatin architecture and remodelling at promoters. Nature *461*, 193–198.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol. Cell *49*, 825–837.

Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. Curr. Opin. Genet. Dev. *14*, 155–164.

Carpenter, A.E., Memedula, S., Plutz, M.J., and Belmont, A.S. (2005). Common effects of

acidic activators on large-scale chromatin structure and transcription. Mol. Cell. Biol. *25*, 958–968.

Chambeyron, S., and Bickmore, W.A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev. *18*, 1119–1130.

Chambeyron, S., Da Silva, N.R., Lawson, K.A., and Bickmore, W.A. (2005). Nuclear re-organisation of the Hoxb complex during mouse embryonic development. Development *132*, 2215–2223.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. Cell *155*, 1479–1491.

Chen, J., Zhang, Z., Li, L., Chen, B.-C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., et al. (2014). Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. Cell *156*, 1274–1285.

Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. *23*, 1163–1171.

Chubb, J.R., Boyle, S., Perry, P., and Bickmore, W.A. (2002). Chromatin motion is constrained by association with nuclear compartments in human cells. Curr. Biol. *12*, 439–445.

Cong, L., Zhou, R., Kuo, Y., Cunniff, M., and Zhang, F. (2012). Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nat. Commun. *3*, 968.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Conrad, S.E., and Botchan, M.R. (1982). Isolation and Characterization of Human DNA

Fragments with Nucleotide Sequence Homologies with the Simian Virus 40 Regulatory Region. Mol. Cell. Biol. *2*.

Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., and Bickmore, W.A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. J. Cell Biol. *145*, 1119–1131.

Davies, J.O.J., Telenius, J.M., McGowan, S.J., Roberts, N.A., Taylor, S., Higgs, D.R., and Hughes, J.R. (2015). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nat. Methods *13*, 74.

Dekker, J., Rippe, K., Dekker, M., Kleckner, N., Woodcock, C.L., Dimitrov, S., Andrulis, E.D., Neiman, A.M., Zappulla, D.C., Sternglanz, R., et al. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell *149*, 1233–1244.

Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A., et al. (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. Cell *158*, 849–860.

Dessaud, E., McMahon, A.P., and Briscoe, J. (2008). Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. Development *135*, 2489–2503.

Ding, Q., Lee, Y.-K., Schaefer, E.A.K., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T., et al. (2013). A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell *12*, 238–251.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during

stem cell differentiation. Nature *518*, 331–336.

Dorman, C.J. (1991). DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria. Infect. Immun. *59*, 745–749.

Dostie, J., and Bickmore, W.A. (2012). Chromosome organization in the nucleus – charting new territory across the Hi-Cs. Curr. Opin. Genet. Dev. *22*, 125–131.

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. Genome Res. *16*, 1299–1309.

Dubourg, C., Bendavid, C., Pasquier, L., Henry, C., Odent, S., and David, V. (2007). Holoprosencephaly. Orphanet J. Rare Dis. *2*, 8.

Echelard, Y., Epstein, D.J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J.A., and McMahon, A.P. (1993). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. Cell *75*, 1417–1430.

van den Engh, G., Sachs, R., and Trask, B.J. (1992). Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. Science *257*, 1410–1412.

Epstein, D.J., McMahon, A.P., and Joyner, A.L. (1999). Regionalization of Sonic hedgehog transcription along the anteroposterior axis of the mouse central nervous system is regulated by Hnf3-dependent and -independent mechanisms. Development *126*, 281–292.

Eskeland, R., Freyer, E., Leeb, M., Wutz, A., and Bickmore, W.A. (2010a). Histone acetylation and the maintenance of chromatin compaction by Polycomb repressive complexes. Cold Spring Harb. Symp. Quant. Biol. *75*, 71–78.

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A., et al. (2010b). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. Mol. Cell *38*, 452–464.

Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., and Skoultchi, A.I. (2005). Histone H1 depletion in mammals alters global

chromatin structure but causes specific changes in gene regulation. Cell *123*, 1199–1212.

Farrar, D., Rai, S., Chernukhin, I., Jagodic, M., Ito, Y., Yammine, S., Ohlsson, R., Murrell, A., and Klenova, E. (2010). Mutational analysis of the poly(ADP-ribosyl)ation sites of the transcription factor CTCF provides an insight into the mechanism of its regulation by poly(ADP-ribosyl)ation. Mol. Cell. Biol. *30*, 1199–1216.

Farrar, D., Chernukhin, I., and Klenova, E. (2011). Generation of poly(ADP-ribosyl)ation deficient mutants of the transcription factor, CTCF. Methods Mol. Biol. *780*, 293–312.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., et al. (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212–224.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574–1577.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell.

Furniss, D., Lettice, L.A., Taylor, I.B., Critchley, P.S., Giele, H., Hill, R.E., and Wilkie, A.O.M. (2008). A variant in the sonic hedgehog regulatory sequence (ZRS) is associated with triphalangeal thumb and deregulates expression in the developing limb. Hum. Mol. Genet. *17*, 2417–2423.

Geng, X., Speirs, C., Lagutin, O., Inbal, A., Liu, W., Solnica-Krezel, L., Jeong, Y., Epstein, D.J., and Oliver, G. (2008). Haploinsufficiency of Six3 fails to activate Sonic hedgehog expression in the ventral forebrain and causes holoprosencephaly. Dev. Cell *15*, 236–247.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature *512*, 96–100.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell *154*, 442–451.

Giorgetti, L., Piolot, T., and Heard, E. (2015). High-resolution 3D DNA FISH using plasmid probes and computational correction of optical aberrations to study chromatin structure at the sub-megabase scale. Methods Mol. Biol. *1262*, 37–53.

Gómez-Marín, C., Tena, J.J., Acemel, R.D., López-Mayorga, M., Naranjo, S., de la Calle-Mustienes, E., Maeso, I., Beccari, L., Aneas, I., Vielmas, E., et al. (2015). Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. Proc. Natl. Acad. Sci. *112*, 7542–7547.

Gottschalk, A.J., Timinszky, G., Kong, S.E., Jin, J., Cai, Y., Swanson, S.K., Washburn, M.P., Florens, L., Ladurner, A.G., Conaway, J.W., et al. (2009). Poly(ADP-ribosyl)ation directs recruitment and activation of an ATP-dependent chromatin remodeler. Proc. Natl. Acad. Sci. U. S. A. *106*, 13770–13774.

Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., and Fraser, P. (2000). Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. Mol. Cell *5*, 377–386.

Gross, D.S., and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. Annu. Rev. Biochem. *57*, 159–197.

Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kollias, G. (1987). Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell *51*, 975–985.

Gruss, P., Dhar, R., and Khoury, G. (1981). Simian virus 40 tandem repeated sequences as an element of the early promoter. Biochemistry *78*, 943–947.

Guastafierro, T., Catizone, A., Calabrese, R., Zampieri, M., Martella, O., Bacalini, M.G., Reale, A., Di Girolamo, M., Miccheli, M., Farrar, D., et al. (2013). ADP-ribose polymer depletion leads to nuclear Ctcf re-localization and chromatin rearrangement(1). Biochem. J. *449*, 623–630.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900–910.

Gustafsson, M.G. (2000). Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. J. Microsc. *198*, 82–87.

Gustafsson, M.G.L., Shao, L., Carlton, P.M., Wang, C.J.R., Golubovskaya, I.N., Cande, W.Z., Agard, D.A., Sedat, J.W., Agard, D.A., Hiraoka, Y., et al. (2008). Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. Biophys. J. *94*, 4957–4970.

Hassa, P.O., Covic, M., Hasan, S., Imhof, R., and Hottiger, M.O. (2001). The enzymatic and DNA binding activity of PARP-1 are not required for NF-kappa B coactivator function. J. Biol. Chem. *276*, 45588–45597.

Hassa, P.O., Haenni, S.S., Buerki, C., Meier, N.I., Lane, W.S., Owen, H., Gersbach, M., Imhof, R., and Hottiger, M.O. (2005). Acetylation of Poly(ADP-ribose) Polymerase-1 by p300/CREB-binding Protein Regulates Coactivation of NF-κB-dependent Transcription. J. Biol. Chem. *280*, 40450–40464.

Hatzis, P., and Talianidis, I. (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. Mol. Cell *10*, 1467–1477.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., et al. (2010). Distinct Epigenomic Landscapes of Pluripotent and Lineage-Committed Human Cells. Cell Stem Cell *6*, 479–491.

Hay, D., Hughes, J.R., Babbs, C., Davies, J.O.J., Graham, B.J., Hanssen, L.L.P., Kassouf, M.T., Oudelaar, A.M., Sharpe, J.A., Suciu, M.C., et al. (2016). Genetic dissection of the α-globin super-enhancer in vivo. Nat. Genet.

Hayashi, K., Tanaka, M., Shimada, T., Miwa, M., and Sugimura, T. (1983). Size and shape of poly(ADP-ribose): examination by gel filtration, gel electrophoresis and electron microscopy. Biochem. Biophys. Res. Commun. *112*, 102–107.

Hecksher-Sørensen, J., Hill, R.E., and Lettice, L. (1998). Double labeling for whole-mount in situ hybridization in mouse. Biotechniques *24*, 914–916, 918.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin

signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Hernandez-Munoz, I., Taghavi, P., Kuijl, C., Neefjes, J., and van Lohuizen, M. (2005). Association of BMI1 with Polycomb Bodies Is Dynamic and Requires PRC2/EZH2 and the Maintenance DNA Methyltransferase DNMT1. Mol. Cell. Biol. *25*, 11047–11058.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat. Genet. *46*, 205–212.

Hutchison, N., and Weintraub, H. (1985). Localization of DNAase I-sensitive sequences to specific regions of interphase nuclei. Cell *43*, 471–482.

Illingworth, R.S., Moffat, M., Mann, A.R., Read, D., Hunter, C.J., Pradeepa, M.M., Adams, I.R., and Bickmore, W.A. (2015). The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. Genes Dev. *29*, 1897–1902.

Inoue, F., Kircher, M., Martin, B., Cooper, G.M., Witten, D.M., McManus, M.T., Ahituv, N., and Shendure, J. (2016). A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity (Cold Spring Harbor Labs Journals).

Izeddin, I., Récamier, V., Bosanac, L., Cissé, I.I., Boudarene, L., Dugast-Darzacq, C., Proux, F., Bénichou, O., Voituriez, R., Bensaude, O., et al. (2014). Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. Elife *3*.

Jeong, Y. (2003). Distinct regulators of Shh transcription in the floor plate and notochord

indicate separate origins for these tissues in the mouse node. Development *130*, 3891–3902.

Jeong, Y. (2006). A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. Development *133*, 761–772.

Jeong, Y., Leskow, F.C., El-Jaick, K., Roessler, E., Muenke, M., Yocum, A., Dubourg, C., Li, X., Geng, X., Oliver, G., et al. (2008). Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. Nat. Genet. *40*, 1348–1353.

Jeong, Y., Dolson, D.K., Waclaw, R.R., Matise, M.P., Sussel, L., Campbell, K., Kaestner, K.H., and Epstein, D.J. (2011). Spatial and temporal requirements for sonic hedgehog in the regulation of thalamic interneuron identity. Development *138*, 531–541.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290–294.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Kim, A., and Dean, A. (2004). Developmental stage differences in chromatin subdomains of the beta-globin locus. Proc. Natl. Acad. Sci. U. S. A. *101*, 7028–7033.

Kim, M.Y., Mauro, S., Gévry, N., Lis, J.T., and Kraus, W.L. (2004). NAD+-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1. Cell *119*, 803–814.

Kim, Y.J., Björklund, S., Li, Y., Sayre, M.H., and Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell *77*, 599–608.

Kind, J., and van Steensel, B. (2010). Genome–nuclear lamina interactions and gene regulation. Curr. Opin. Cell Biol. *22*, 320–325.

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C., et al. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nat. Struct. Mol. Biol. *18*, 956–963.

Kolesky, S.E., Ouhammouch, M., and Geiduschek, E.P. (2002). The mechanism of transcriptional activation by the topologically DNA-linked sliding clamp of bacteriophage T4. J. Mol. Biol. *321*, 767–784.

Kouzarides, T. (2007). SnapShot: Histone-modifying enzymes. Cell *131*, 822.

Kraus, W.L., and Lis, J.T. (2003). PARP goes transcription. Cell *113*, 677–683.

Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. Mol. Cell *44*, 410–423.

Krishnakumar, R., and Kraus, W.L. (2010). PARP-1 regulates chromatin structure and transcription through a KDM5B-dependent pathway. Mol. Cell *39*, 736–749.

Krishnakumar, R., Gamble, M.J., Frizzell, K.M., Berrocal, J.G., Kininis, M., and Kraus, W.L. (2008). Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. Science *319*, 819–821.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. *4*, e1000242.

Kundu, T.K., Palhan, V.B., Wang, Z., An, W., Cole, P.A., and Roeder, R.G. (2000). Activator-Dependent Transcription from Chromatin In Vitro Involving Targeted Histone Acetylation by p300. Mol. Cell *6*, 551–561.

Kvon, E.Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature.

de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and

their regulatory landscapes. Nature *502*, 499–506.

Lebedev, D.V., Filatov, M.V., Kuklin, A.I., Islamov, A.K., Kentzinger, E., Pantina, R., Toperverg, B.P., and Isaev-Ivanov, V.V. (2005). Fractal nature of chromatin organization in interphase chicken erythrocyte nuclei: DNA structure exhibits biphasic fractal properties.

Lettice, L.A., Horikoshi, T., Heaney, S.J.H., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N., et al. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc. Natl. Acad. Sci. *99*, 7548–7553.

Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. *12*, 1725–1735.

Lettice, L.A., Hill, A.E., Devenney, P.S., and Hill, R.E. (2008). Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. Hum. Mol. Genet. *17*, 978–985.

Lettice, L.A., Williamson, I., Devenney, P.S., Kilanowski, F., Dorin, J., and Hill, R.E. (2014). Development of five digits is controlled by a bipartite long-range cis-regulator. Development *141*, 1715–1725.

Li, Q. (2002). Locus control regions. Blood *100*, 3077–3086.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707–719.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Loots, G.G., Ovcharenko, I., Pachter, L., Dubchak, I., and Rubin, E.M. (2002). rVista for Comparative Sequence-Based Discovery of Functional Transcription Factor Binding Sites. Genome Res. *12*, 832–839.

Lucas, J.S., Zhang, Y., Dudko, O.K., and Murre, C. (2014). 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. Cell *158*, 339–352.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.

Maas, S.A., Suzuki, T., and Fallon, J.F. (2011). Identification of spontaneous mutations within the long-range limb-specific Sonic hedgehog enhancer (ZRS) that alter Sonic hedgehog expression in the chicken limb mutants oligozeugodactyly and silkie breed. Dev. Dyn. *240*, 1212–1222.

Mahmoudi, T. (2002). GAGA can mediate enhancer function in trans by linking two separate DNA molecules. EMBO J. *21*, 1775–1781.

Mahy, N.L., Perry, P.E., and Bickmore, W.A. (2002). Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. J. Cell Biol. *159*, 753–763.

Markaki, Y., Smeets, D., Fiedler, S., Schmid, V.J., Schermelleh, L., Cremer, T., and Cremer, M. (2012). The potential of 3D-FISH and super-resolution structured illumination microscopy for studies of 3D nuclear architecture. BioEssays *34*, 412–426.

Masternak, K., Peyraud, N., Krawczyk, M., Barras, E., and Reith, W. (2003). Chromatin remodeling and extragenic transcription at the MHC class II locus control region. Nat. Immunol. *4*, 132–137.

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.-Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *44*, D110–D115.

Matsumoto, K., and Hirose, S. (2004). Visualization of unconstrained negative supercoils of DNA on polytene chromosomes of Drosophila. J. Cell Sci. *117*, 3797–3805.

McMahon, A.P., Ingham, P.W., and Tabin, C.J. (2003). 1 Developmental roles and clinical

significance of Hedgehog signaling. Curr. Top. Dev. Biol. *53*, 1–114.

Mellon, P., Parker, V., Gluzman, Y., and Maniatis, T. (1981). Identification of DNA sequences required for transcription of the human α1-globin gene in a new SV40 host-vector system. Cell *27*, 279–288.

Merkenschlager, M., and Odom, D.T. (2013). CTCF and Cohesin: Linking Gene Regulatory Elements with Their Targets. Cell *152*, 1285–1297.

Mirny, L.A. (2011). The fractal globule as a model of chromatin architecture in the cell. Chromosome Res. *19*, 37–51.

Misteli, T., Gunjan, A., Hock, R., Bustin, M., and Brown, D.T. (2000). Dynamic binding of histone H1 to chromatin in living cells. Nature *408*, 877–881.

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., and Duboule, D. (2011). A Regulatory Archipelago Controls Hox Genes Transcription in Digits. Cell *147*, 1132–1145.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, P., and Chambon (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res.

Morey, C., Da Silva, N.R., Perry, P., and Bickmore, W.A. (2007). Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation. Development *134*, 909–919.

Morey, C., Kress, C., and Bickmore, W.A. (2009). Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression. Genome Res. *19*, 1184–1194.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science *326*, 1501.

Nalabothula, N., Al-jumaily, T., Eteleeb, A.M., Flight, R.M., Xiaorong, S., Moseley, H., Rouchka, E.C., and Fondufe-Mittendorf, Y.N. (2015). Genome-Wide Profiling of PARP1 Reveals an Interplay with Gene Regulatory Regions and DNA Methylation. PLoS One *10*,

e0135410.

Naughton, C., Avlonitis, N., Corless, S., Prendergast, J.G., Mati, I.K., Eijk, P.P., Cockroft, S.L., Bradley, M., Ylstra, B., and Gilbert, N. (2013). Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. Nat. Struct. Mol. Biol. *20*, 387–395.

Nichols, M.H., and Corces, V.G. (2015). A CTCF Code for 3D Genome Architecture. Cell *162*, 703–705.

Noonan, J.P., and McCallion, A.S. (2010). Genomics of Long-Range Regulatory Elements. Annu. Rev. Genomics Hum. Genet. *11*, 1–23.

Noordermeer, D., Branco, M.R., Splinter, E., Klous, P., van IJcken, W., Swagemakers, S., Koutsourakis, M., van der Spek, P., Pombo, A., and de Laat, W. (2008). Transcription and Chromatin Organization of a Housekeeping Gene Cluster Containing an Integrated β-Globin Locus Control Region. PLoS Genet. *4*, e1000016.

Noordermeer, D., de Wit, E., Klous, P., van de Werken, H., Simonis, M., Lopez-Jones, M., Eussen, B., de Klein, A., Singer, R.H., and de Laat, W. (2011). Variegated gene expression caused by cell-specific long-range DNA interactions. Nat. Cell Biol. *13*, 944–951.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nord, A.S., Pattabiraman, K., Visel, A., and Rubenstein, J.L.R. (2015). Genomic perspectives of transcriptional regulation in forebrain development. Neuron *85*, 27–47.

Normanno, D., Boudarène, L., Dugast-Darzacq, C., Chen, J., Richter, C., Proux, F., Bénichou, O., Voituriez, R., Darzacq, X., and Dahan, M. (2015). Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. Nat. Commun. *6*, 7357.

Ogino, H., Nozaki, T., Gunji, A., Maeda, M., Suzuki, H., Ohta, T., Murakami, Y., Nakagama, H., Sugimura, T., and Masutani, M. (2007). Loss of Parp-1 affects gene expression profile in a genome-wide manner in ES cells and liver cells. BMC Genomics *8*,

41.

Ong, C.-T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. *15*, 234–246.

Palstra, R.-J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. Nat. Genet. *35*, 190–194.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell *1*, 299–312.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell *132*, 422–433.

Pavri, R., Lewis, B., Kim, T.-K., Dilworth, F.J., Erdjument-Bromage, H., Tempst, P., de Murcia, G., Evans, R., Chambon, P., and Reinberg, D. (2005). PARP-1 determines specificity in a retinoid signaling pathway via direct modulation of mediator. Mol. Cell *18*, 83–96.

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nat. Biotechnol. *24*, 79–88.

Petesch, S.J., and Lis, J.T. (2008). Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at Hsp70 loci. Cell *134*, 74–84.

Petesch, S.J., and Lis, J.T. (2012). Activator-induced spread of poly(ADP-ribose) polymerase promotes nucleosome loss at Hsp70. Mol. Cell *45*, 64–74.

Pirrotta, V., and Li, H.-B. (2012). A view of nuclear Polycomb bodies. Curr. Opin. Genet. Dev. *22*, 101–109.

Poirier, G.G., de Murcia, G., Jongstra-Bilen, J., Niedergang, C., and Mandel, P. (1982). Poly(ADP-ribosyl)ation of polynucleosomes causes relaxation of chromatin structure. Proc.

Natl. Acad. Sci. U. S. A. *79*, 3423–3427.

Pollard, S.M., Benchoua, A., and Lowell, S. (2006). Neural stem cells, neurons, and glia. Methods Enzymol. *418*, 151–169.

Pradeepa, M.M., Grimes, G.R., Kumar, Y., Olley, G., Taylor, G.C.A., Schneider, R., and Bickmore, W.A. (2016). Histone H3 globular domain acetylation identifies a new class of enhancers. Nat. Genet.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569–577.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M., et al. (2011). Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. Cell *147*, 1628–1639.

Ran, F.A., Hsu, P.D., Lin, C.-Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., et al. (2013a). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380–1389.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Ravi, V., Bhatia, S., Gautier, P., Loosli, F., Tay, B.-H., Tay, A., Murdoch, E., Coutinho, P., van Heyningen, V., Brenner, S., et al. (2013). Sequencing of Pax6 loci from the elephant shark reveals a family of Pax6 genes in vertebrate genomes, forged by ancient duplications and divergences. PLoS Genet. *9*, e1003177.

Riddle, R.D., Johnson, R.L., Laufer, E., and Tabin, C. (1993). Sonic hedgehog mediates the

polarizing activity of the ZPA. Cell *75*, 1401–1416.

Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu. Rev. Genet. *38*, 413–443.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Roessler, E., Belloni, E., Gaudenz, K., Jay, P., Berta, P., Scherer, S.W., Tsui, L.C., and Muenke, M. (1996). Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. Nat. Genet. *14*, 357–360.

Roh, T., Wei, G., Farrell, C.M., and Zhao, K. (2007). Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. Genome Res. *17*, 74–81.

Roh, T.-Y., Cuddapah, S., and Zhao, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev. *19*, 542–552.

Ronshaugen, M., and Levine, M. (2004). Visualization of trans-Homolog Enhancer-Promoter Interactions at the Abd-B Hox Locus in the Drosophila Embryo. Dev. Cell *7*, 925–932.

Sachs, R.K., van den Engh, G., Trask, B., Yokota, H., and Hearst, J.E. (1995). A random-walk/giant-loop model for interphase chromosomes. Proc. Natl. Acad. Sci. U. S. A. *92*, 2710–2714.

Sagai, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. Development *132*, 797–803.

Sagai, T., Masuya, H., Tamura, M., Shimizu, K., Yada, Y., Wakana, S., Gondo, Y., Noda, T., and Shiroishi, T. (2004). Phylogenetic conservation of a limb-specific, cis -acting regulator of Sonic hedgehog (Shh). Mamm. Genome *15*, 23–34.

Sagai, T., Amano, T., Tamura, M., Mizushina, Y., Sumiyama, K., and Shiroishi, T. (2009). A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. Development *136*, 1665–1674.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113.

Schirm, S., Jiricny, J., and Schaffner, W. (1987). The SV40 enhancer can be dissected into multiple segments, each with a different cell type specificity. Genes Dev *1*, 65–74.

Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.-M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S.W., et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S., et al. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat. Genet. *42*, 53–61.

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.-M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M., et al. (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. Nat. Genet. *47*, 1179–1186.

Scott, K.C., Taubman, A.D., and Geyer, P.K. (1999). Enhancer blocking by the Drosophila gypsy insulator depends upon insulator anatomy and enhancer strength. Genetics *153*, 787–798.

Seitan, V.C., Faure, A.J., Zhan, Y., McCord, R.P., Lajoie, B.R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A.G., et al. (2013). Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res. *23*, 2066–2077.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. Cell *148*, 458–472.

Shopland, L.S., Lynch, C.R., Peterson, K.A., Thornton, K., Kepper, N., Hase, J. von, Stein, S., Vincent, S., Molloy, K.R., Kreth, G., et al. (2006). Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary

genomic sequence. J. Cell Biol. *174*, 27–38.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). Nat. Genet. *38*, 1348–1354.

Sofueva, S., Yaffe, E., Chan, W.-C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. EMBO J. *32*, 3119–3129.

Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. Nat. Rev. Cancer *6*, 846–856.

Spicuglia, S., Kumar, S., Yeh, J.-H., Vachez, E., Chasson, L., Gorbatch, S., Cautres, J., and Ferrier, P. (2002). Promoter activation by enhancer-dependent and -independent loading of activator and coactivator complexes. Mol. Cell *10*, 1479–1487.

Spitz, F., Gonzalez, F., and Duboule, D. (2003). A Global Control Region Defines a Chromosomal Regulatory Landscape Containing the HoxD Cluster. Cell *113*, 405–417.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. *20*, 267–273.

Su, W., Jackson, S., Tjian, R., and Echols, H. (1991). DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes Dev. *5*, 820–826.

Sutherland, H., and Bickmore, W.A. (2009). Transcription factories: gene expression in unions? Nat. Rev. Genet. *10*, 457–466.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat. Rev. Genet. *9*, 465–476.

Symmons, O., and Spitz, F. (2013). From remote enhancers to gene regulation: charting the genome's regulatory landscapes. Philos. Trans. R. Soc. B Biol. Sci. *368*, 20120358–20120358.

Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory

domains. Genome Res. *24*, 390–400.

Taatjes, D.J. (2010). The human Mediator complex: a versatile, genome-wide regulator of transcription. Trends Biochem. Sci. *35*, 315–322.

Tanabe, H., Müller, S., Neusser, M., von Hase, J., Calcagno, E., Cremer, M., Solovei, I., Cremer, C., and Cremer, T. (2002). Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. Proc. Natl. Acad. Sci. U. S. A. *99*, 4424–4429.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. Cell *159*, 635–646.

Taylor, G.C.A., Eskeland, R., Hekimoglu-Balkan, B., Pradeepa, M.M., and Bickmore, W.A. (2013). H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. Genome Res. *23*, 2053–2065.

Therizols, P., Illingworth, R.S., Courilleau, C., Boyle, S., Wood, A.J., and Bickmore, W.A. (2014). Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. Science *346*, 1238–1242.

Thoma, F., and Koller, T. (1977). Influence of histone H1 on chromatin structure. Cell *12*, 101–107.

Timinszky, G., Till, S., Hassa, P.O., Hothorn, M., Kustatscher, G., Nijmeijer, B., Colombelli, J., Altmeyer, M., Stelzer, E.H.K., Scheffzek, K., et al. (2009). A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. Nat. Struct. Mol. Biol. *16*, 923–929.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol. Cell *10*, 1453–1465.

Tóth, K.F., Knoch, T.A., Wachsmuth, M., Frank-Stöhr, M., Stöhr, M., Bacher, C.P., Müller, G., and Rippe, K. (2004). Trichostatin A-induced histone acetylation causes decondensation of interphase chromatin. J. Cell Sci. *117*, 4277–4287.

Tsujimura, T., Klein, F.A., Langenfeld, K., Glaser, J., Huber, W., Spitz, F., Visel, A., Rubin, E., Rubin, E., Pennacchio, L., et al. (2015). A Discrete Transition Zone Organizes the Topological and Regulatory Autonomy of the Adjacent Tfap2c and Bmp7 Genes. PLoS Genet. *11*, e1004897.

Tsukiji, N., Amano, T., and Shiroishi, T. (2014). A novel regulatory element for Shh expression in the lung and gut of mouse embryos. Mech. Dev. *131*, 127–136.

Tulin, A., and Spradling, A. (2003). Chromatin loosening by poly(ADP)-ribose polymerase (PARP) at Drosophila puff loci. Science *299*, 560–562.

Tulin, A., Stewart, D., and Spradling, A.C. (2002). The Drosophila heterochromatic gene encoding poly(ADP-ribose) polymerase (PARP) is required to modulate chromatin structure during development. Genes Dev. *16*, 2108–2119.

Tulin, A., Chinenov, Y., and Spradling, A. (2003). Regulation of chromatin structure and gene activity by poly(ADP-ribose) polymerases. Curr. Top. Dev. Biol. *56*, 55–83.

Tumbar, T., Sudlow, G., and Belmont, A.S. (1999). Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. J. Cell Biol. *145*, 1341–1354.

Turner, B.M., Birley, A.J., and Lavender, J. (1992). Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. Cell *69*, 375–384.

Vernimmen, D., and Bickmore, W.A. (2015). The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. Trends Genet. *31*, 696–708.

Vernimmen, D., De Gobbi, M., Sloane-Stanley, J.A., Wood, W.G., and Higgs, D.R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. EMBO J. *26*, 2041–2051.

Vernimmen, D., Lynch, M.D., De Gobbi, M., Garrick, D., Sharpe, J.A., Sloane-Stanley, J.A., Smith, A.J.H., and Higgs, D.R. (2011). Polycomb eviction as a new distant enhancer function. Genes Dev. *25*, 1583–1588.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry,

M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854–858.

Vojnic, E., Mourão, A., Seizl, M., Simon, B., Wenzeck, L., Larivière, L., Baumli, S., Baumgart, K., Meisterernst, M., Sattler, M., et al. (2011). Structure and VP16 binding of the Mediator Med25 activator interaction domain. Nat. Struct. Mol. Biol. *18*, 404–409.

Volpi, E. V, Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R.D., Goldsworthy, M., Powis, S.H., Ragoussis, J., et al. (2000). Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. J. Cell Sci. *113* , 1565–1576.

Wallis, D.E., Roessler, E., Hehr, U., Nanni, L., Wiltshire, T., Richieri-Costa, A., Gillessen-Kaesbach, G., Zackai, E.H., Rommens, J., and Muenke, M. (1999). Mutations in the homeodomain of the human SIX3 gene cause holoprosencephaly. Nat. Genet. *22*, 196–198.

Wang, Q., Carroll, J.S., and Brown, M. (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. Mol. Cell *19*, 631–642.

Wang, S., Su, J.-H., Beliveau, B.J., Bintu, B., Moffitt, J.R., Wu, C.-T., and Zhuang, X. (2016). Spatial organization of chromatin domains and compartments in single chromosomes. Science.

Williamson, I., Eskeland, R., Lettice, L.A., Hill, A.E., Boyle, S., Grimes, G.R., Hill, R.E., and Bickmore, W.A. (2012). Anterior-posterior differences in HoxD chromatin topology in limb development. Development *139*, 3157–3167.

Williamson, I., Berlivet, S., Eskeland, R., Boyle, S., Illingworth, R.S., Paquette, D., Dostie, J., and Bickmore, W.A. (2014). Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. Genes Dev. *28*, 2778–2791.

Williamson, I., Lettice, L.A., Hill, R.E., and Bickmore, W.A. (2016). Shh and ZRS enhancer co-localisation is specific to the zone of polarizing activity. Development dev.139188.

de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. Genes Dev. *26*, 11–24.

Woodcock, C.L., Skoultchi, A.I., and Fan, Y. (2006). Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. Chromosome Res. *14*, 17–25.

Woringer, M., Darzacq, X., and Izeddin, I. (2014). Geometry of the nucleus: a perspective on gene expression regulation. Curr. Opin. Chem. Biol. *20*, 112–119.

Wu, S.M., Choo, A.B.H., Yap, M.G.S., and Chan, K.K.-K. (2010). Role of Sonic hedgehog signaling and the expression of its components in human embryonic stem cells. Stem Cell Res. *4*, 38–49.

Xiao, B., Freedman, B.S., Miller, K.E., Heald, R., and Marko, J.F. (2012). Histone H1 compacts DNA under force and during chromatin assembly. Mol. Biol. Cell *23*, 4864–4871.

Xu, Z., Wei, G., Chepelev, I., Zhao, K., and Felsenfeld, G. (2011). Mapping of INS promoter interactions reveals its role in long-range regulation of SYT8 transcription. Nat. Struct. Mol. Biol. *18*, 372–378.

Yaffe, E., and Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat. Genet. *43*, 1059–1065.

Yao, Y., Minor, P.J., Zhao, Y.-T., Jeong, Y., Pani, A.M., King, A.N., Symmons, O., Gan, L., Cardoso, W. V, Spitz, F., et al. (2016). Cis-regulatory architecture of a brain signaling center predates the origin of chordates. Nat. Genet.

Yin, J., Wang, G., Adelman, K., Lis, J.T., Akoulitchev, S., Chuikov, S., Reinberg, D., Apostolou, E., Ferrari, F., Walsh, R.M., et al. (2014). The Mediator complex: a master coordinator of transcription and cell lineage development. Development *141*, 977–987.

Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat. Biotechnol. *21*, 183–186.

Yokota, H., Singer, M.J., van den Engh, G.J., and Trask, B.J. (1997). Regional differences in the compaction of chromatin in human G0/G1 interphase nuclei. Chromosome Res. *5*, 157–166.

Yu, W., Ginjala, V., Pant, V., Chernukhin, I., Whitehead, J., Docquier, F., Farrar, D., Tavoosidana, G., Mukhopadhyay, R., Kanduri, C., et al. (2004). Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. Nat. Genet. *36*, 1105–1110.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., and Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat. Biotechnol. *29*, 149–153.

Zhao, H., and Dean, A. (2004). An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. Nucleic Acids Res. *32*, 4903–4919.

Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell *152*, 642–654.

Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M., and Tuan, D. (2007). A facilitated tracking and transcription mechanism of long-range enhancer function. Nucleic Acids Res. *35*, 5532–5544.

Zorn, C., Cremer, T., Cremer, C., and Zimmer, J. (1976). Laser UV microirradiation of interphase nuclei and post-treatment with caffeine. A new approach to establish the arrangement of interphase chromosomes. Hum. Genet. *35*, 83–89.

Zorn, C., Cremer, C., Cremer, T., and Zimmer, J. (1979). Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. Distribution in interphase and metaphase. Exp. Cell Res. *124*, 111–119.

Zuin, J., Dixon, J.R., van der Reijden, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van IJcken, W.F.J., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. U. S. A. *111*, 996–1001.